Nonlinear signal processing by noisy spiking neurons

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)
im Fach Physik
Spezialisierung: Theoretische Physik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von Herrn
M. Sc. Sergej Olegovic Voronenko

Präsidentin der Humboldt-Universität zu Berlin:
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät:
Prof. Dr. Elmar Kulke

Gutachter:
1. Prof. Dr. Benjamin Lindner (Humboldt-Universität zu Berlin)
2. Prof. Dr. Igor Sokolov (Humboldt-Universität zu Berlin)
3. Prof. Dr. Stefan Rotter (Albert-Ludwigs-Universität Freiburg)

Die Arbeit wurde eingereicht am: 28. September 2017
Tag der mündlichen Prüfung: 23. Dezember 2017

Abstract

Neurons are excitable cells which communicate with each other via electrical signals. In general, these signals are processed by the Neurons in a nonlinear fashion, the exact mathematical description of which is still an open problem in neuroscience. In this thesis, the broad topic of nonlinear signal processing is approached from two directions. The first part of the thesis is devoted to the question how input signals modulate the neural response. The second part of the thesis is concerned with the nonlinear reconstruction of input signals from the neural output and with the estimation of the amount of the transmitted information.

The nonlinear modulation of the neural response is studied by extending the linear response theory. For the analytically tractable leaky integrate-and-fire model, we derive the second-order response function that allows to predict the time-dependent firing rate in response to general time-dependent signals, which is not only a single-neuron statistics which is often measured in experiments, but can also be regarded as the instantaneous population rate of a group of uncoupled neurons subject to the same external drive. The weakly nonlinear theory reveals for the leaky integrate-and-fire model several interesting features. For cosine signals, the theory successfully describes the excitation of higher harmonics, a frequency-dependent change of the time-averaged firing rate, and a strong nonlinear interaction of multiple input signals. The weakly nonlinear regime emerges as a transition between the regime of linear encoding and a highly nonlinear regime in which the neuron exhibits mode-locking behavior. The transition emerges not only for increasing signal amplitude, as suggested by the Volterra expansion of the firing rate, but also for decreasing strength of the intrinsic noise. Beyond the analytically tractable leaky integrate-and-fire model with white Gaussian current noise, we also measure numerically the second-order response function for a biophysically more realistic spiking neuron model with channel noise. The similarity of the weakly nonlinear response for the two models suggests that the discussed features do not hinge on the particularities of the models but are of a more general nature.

The nonlinear reconstruction of input signals and the nonlinear estimation of the information content of the neural spike count is studied by extending the linear lower bound for the mutual information. For the special case of a static input signal, we derive a quadratic lower bound which can be used as a simple estimate for the mutual information. Our result is expressed in terms of linear and nonlinear correlation coefficients which can be computed for any signal-processing system with time-independent input and output. From a decoding perspective, the quadratic lower bound corresponds to the minimal information that is recovered from the output via a quadratic reconstruction. We demonstrate that the quadratic signal reconstruction outperforms the linear theory if the mean of the output variable exhibits a nonlinear signal dependence. More interestingly, however, the quadratic lower bound can be used as an estimate of the mutual information for systems which exhibit a signal-dependent variance and for which the linear lower bound is always zero. For two different neuron models we demonstrate that a quadratic reconstruction of the input signal can outperform the linear reconstruction significantly. This finding implies that higher-order correlations between signal and spike count carry information which is not accounted for by the linear correlations. For the estimation of the mutual information between signal and spike count we find that the linear and

the quadratic lower bounds tightly approximate the mutual information for weak signals. For stronger signals, however, the quadratic lower bound provides an accurate estimate of the mutual information for a larger range of signal strengths than the linear lower bound.

The results of this thesis demonstrate how existing linear theories can be extended to capture nonlinear contributions of the signal to the neural response or to incorporate nonlinear correlations into the estimation of the transmitted information. More importantly, however, our analysis demonstrates that these extensions do not merely provide small corrections to the existing linear theories but can account for qualitatively novel effects which are completely missed by the linear theories. These effects include, for example, the excitation of harmonic oscillations in the neural firing rate or the estimation of information for systems with a signal-dependent output variance.

Deutschsprachige Zusammenfassung

Neurone sind anregbare Zellen, die mit Hilfe von elektrischen Signalen miteinander kommunizieren. Im allgemeinen werden eingehende Signale von den Nervenzellen in einer nichtlinearen Art und Weise verarbeitet. Wie diese Verarbeitung in einer umfassenden und exakten Art und Weise mathematisch beschrieben werden kann, ist bis heute nicht geklärt und ist Gegenstand aktueller Forschung. In dieser Arbeit untersuchen wir die nichtlineare Übertragung und Verarbeitung von Signalen durch stochastische Nervenzellen und wenden dabei zwei unterschiedliche Herangehensweisen an. Im ersten Teil der Arbeit befassen wir uns mit der Frage, auf welche Art und Weise ein Signal mit einer bekannten Zeitabhängigkeit die Rate der neuronalen Aktivität beeinflusst. Im zweiten Teil der Arbeit widmen wir uns der Rekonstruktion eingehender Signale aus der durch sie hervorgerufenen neuronalen Aktivität und beschäftigen uns mit der Abschätzung der übertragenen Informationsmenge.

Wir untersuchen die nichtlineare Modulation der neuronalen Aktivitätsrate, indem wir die Theorie der linearen Antwort um die nächsthöhere Ordnung erweitern. Für das Integratorneuron mit Leckstrom (LIF Neuron, eng: leaky integrate-and-fire neuron) präsentieren wir eine analytische Herleitung der Antwortfunktion zweiter Ordnung und stellen eine Theorie der schwach nichtlinearen Antwort auf. Diese Theorie ermöglicht es die nichtlineare Modulation der neuronalen Aktivitätsrate einer einzelnen Zelle zu beschreiben, kann aber auch für die Beschreibung der Aktivität einer Population von ungekoppelten Nervenzellen, die mit einem gemeinsamen Signal getrieben werden, verwendet werden. Für das LIF Neuron zeigt die schwach nichtlineare Theorie die folgenden nichtlinearen Effekte auf: i) Kosinussignale können höhere harmonische Oszillationen der neuronalen Aktivitätsrate anregen. ii) Kosinussignale können die zeitgemittelte Aktivitätsrate in einer nichtlinearen Art und Weise erhöhen oder verringern. iii) Mehrere Kosinussignale mit unterschiedlichen Frequenzen führen zu einer Frequenzmischung in der neuronalen Antwort. Interessanterweise ist der Grad der Ausprägung der schwach nichtlinearen Effekte nicht nur von der Signalamplitude sondern auch von der Stärke des intrinsischen Rauschens abhängig. Die Antwortfunktion zweiter Ordnung kann nicht nur analytisch berechnet, sondern kann auch aus experimentellen oder numerischen Daten bestimmt werden. Für ein biophysikalisch komplexeres Neuronenmodell als das LIF Modell, ergibt eine solche numerische Messung, dass die von uns beschriebenen schwach nichtlinearen Effekte für beide Modelle gleichermaßen auftreten und somit robust sind.

Für die Untersuchung der nichtlinearen Rekonstruktion von Signalen erweitern wir die Theorie der linearen Signalrekonstruktion, indem wir quadratische Beiträge der neuronalen Aktivität bei der Signalrekonstruktion mitberücksichtigen. Für statische Signale leiten wir anschließend eine quadratische untere Schranke her, die eine einfache Abschätzung für die Transinformation zwischen einem Signal und der Aktivität des durch das Signal getriebenen Neurons erlaubt. Während die direkte Messung der Transinformation nur mit großen Datenmengen möglich ist, die für biologische Neuronen manchmal schlichtweg nicht in der erforderlichen Menge verfügbar sind, lässt sich die untere Schranke mit Hilfe nichtlinearer Korrelationskoeffizienten zwischen dem Signal und der neuronalen Antwort ausdrücken und lässt sich für jedes beliebige System, welches eine zeitunabhängige Signalvariable in eine zeitunabhängige Antwortvariable überführt, problemlos auswerten. Die quadratische untere Schranke

stellt dabei nicht nur einen Schätzwert für die Transinformation dar, sondern gibt auch die Menge an Information an, die durch einen optimalen quadratischen Filter aus der neuronalen Aktivität extrahiert werden kann. Für ein Gaußsches Modell, welches eine Signalvariable in eine Antwortvariable überführt, zeigen wir, dass, insbesondere dann wenn der Mittelwert der Ausgangsvariable eine schwach nichtlineare Signalabhängigkeit aufweist, der quadratische Filter signifikant mehr Information aus der Antwortvariablen extrahiert als ein linearer Filter. Der interessantere Fall tritt allerdings ein, wenn der Mittelwert der Antwortvariablen signalunabhängig ist und lediglich die Varianz der Antwort eine Signalabhängigkeit aufweist. In diesem Fall kann die quadratische untere Schranke erfolgreich als Schätzwert für die Transinformation verwendet werden, wohingegen die lineare untere Schranke immer null ist. Für zwei verschiedene Neuronenmodelle zeigen wir, dass die quadratische untere Schranke einen besseren Schätzwert für die Transinformation darstellt als die lineare untere Schranke, da sie für eine größere Spanne an Signalstärken gültig ist. Ferner demonstrieren wir, dass ein quadratischer Filter für bestimmte Signalstärken wesentlich mehr Information aus der neuronalen Aktivität extrahiert als der lineare Filter. Dieses Ergebnis weist darauf hin, dass Korrelationen höherer Ordnung zwischen dem Signal und der neuronalen Aktivität Information über das Signal enthalten, die durch den linearen Korrelationskoeffizienten nicht erfasst wird.

Die Ergebnisse dieser Arbeit demonstrieren, wie die etablierten linearen Theorien, die die Modellierung der neuronalen Aktivitätsrate bzw. die Rekonstruktion von Signalen beschreiben, um Beiträge höherer Ordnung erweitert werden können. Einen wichtigen Beitrag dieser Arbeit stellt allerdings auch die Darstellung der Signifikanz der nichtlinearen Theorien dar. Die nichtlinearen Beiträge erweisen sich nicht nur als schwache Korrekturen zu den etablierten linearen Theorien, sondern beschreiben neuartige Effekte, die durch die linearen Theorien nicht erfasst werden können. Zu diesen Effekten gehört zum Beispiel die Anregung von harmonischen Oszillationen der neuronalen Aktivitätsrate und die Kodierung von Signalen in der signalabhängigen Varianz einer Antwortvariablen.

Contents

List of Symbols and Abbreviations

$\langle . \rangle_\eta$	ensemble average over the random variable η keeping all other random variables fixed
$\langle . \rangle_T$	time average
r_0	unperturbed firing rate (in absence of a signal)
$r(t)$	time-dependent firing rate (modulated by a signal)
$\tilde{x}(\omega) = \int_{-\infty}^{\infty} dt\ e^{i\omega t} x(t)$	Fourier transform of the function or process $x(t)$
$\tilde{x}_T(\omega) = \int_0^T dt\ e^{i\omega t} x(t)$	Fourier transform of the function or process $x(t)$ for finite observation time T
$\tilde{x} * (\omega)$	complex conjugated of $\tilde{x}(\omega)$
$S_{xx}(\omega)$	spike train power spectrum
$S_{ss}(\omega)$	signal power spectrum
$S_{xs}(\omega)$	cross spectrum between spike train and signal
$R(\tau)$	correlation function
$C(\omega)$	coherence function
$K_1(t)$, $K_2(t_1, t_2)$	first- and second-order response kernel
$\chi_1(\omega)$, $\chi_2(\omega_1, \omega_2)$	first- and second-order response function (Fourier transforms of K_1 and K_2)
LIF model	leaky integrate-and-fire neuron model
NaK model	sodium and potassium neuron model

$P(x_j)$	probability that the variable x is in state j
$P(x_j, s_i)$	joint probability that the variables x and s are in states j and i, respectively
$P(x_j \mid s_i)$	conditional probability that the variable x is in state j given that the variable s is in state i
$p(x)$	probability density of the variable x
MI	mutual information
MI_{lb}^{lin}	linear lower bound for the mutual information
MI_{lb}^{quad}	quadratic lower bound for the mutual information
MI_{ub}	upper bound for the mutual information
$\sigma_N = \sqrt{\left\langle \left(N - \langle N \rangle\right)^2 \right\rangle}$	standard deviation of variable N
$\sigma_{N^2} = \sqrt{\left\langle \left(N^2 - \langle N^2 \rangle\right)^2 \right\rangle}$	standard deviation of variable N^2
$\rho_{s,N} = \frac{\left\langle \left(s - \langle s \rangle\right)\left(N - \langle N \rangle\right) \right\rangle}{\sigma_s \sigma_N}$	linear correlation coefficient between variables s and N
$\rho_{s,N^2} = \frac{\left\langle \left(s - \langle s \rangle\right)\left(N^2 - \langle N^2 \rangle\right) \right\rangle}{\sigma_s \sigma_{N^2}}$	correlation coefficient between variables s and N^2
$\rho_{N,N^2} = \frac{\left\langle \left(N - \langle N \rangle\right)\left(N^2 - \langle N^2 \rangle\right) \right\rangle}{\sigma_N \sigma_{N^2}}$	correlation coefficient between variables N and N^2

List of Figures

1 Introduction

Abstract | In this chapter we will introduce the main topic of this thesis, the mathematical tools that we will use for the mathematical analysis, and the neural models that we will employ for the numerical simulations.

In order to survive in an ever changing environment, successfully find food, recognize potential mating partners or escape predators, many organisms developed over the course of evolution an organ that allows them to process sensory stimuli and respond to these stimuli via movement and behavioral responses. This organ is the brain, a structure which consists of many little elements, the nerve cells or neurons, which are interconnected in a complex manner and interact with each other via electrical signals. While in the beginning of the 20th century neuroscience was mostly the realm of clinicians and electrophysiologists, over the years more and more scientific disciplines became involved into the research of brain function [1]. Today, many approaches to study the brain coexist and range from behavioral studies in psychophysics [2], functional magnetic resonance imaging (fMRI) studies of the blood flow in different brain areas [3], and EEG measurements of patients with neural impairments [4] to electrophysiological studies of the activity of single neurons [5]. This large range of different approaches reflects a very interesting feature of the brain that has also a great appeal to theoretical physicists: The properties of very small elements, the neurons, that are only some micrometers in size, determine the emergence of complex phenomena on a macroscopic level like the properties of neural networks or even behavior and cognition. It is this influence of microscopic elements on macroscopic phenomena over multiple orders of magnitude that makes understanding the brain a very difficult task.

In order to one day be able to understand the brain on a macroscopic level, we first have to understand the dynamics of the elements that this complex organ is built of - the neurons. Neurons are electrically active cells, which communicate with each other via voltage pulses, which are also referred to as spikes. How exactly neurons encode incoming signals into outgoing spikes is still not fully understood and the search for this *neural code* remains a hot topic of ongoing research [6, 7]. If we knew what the neural code was, it would enable us to either predict the neural response to an incoming signal (signal encoding) or it would enable us to infer the incoming signal from an observation

of the outgoing spikes (signal decoding). In this thesis, we will try to uncover some aspects of neural signal encoding and decoding by studying mathematically tractable neuron models. Although neurons are in general complex cells which have intricate morphological structures we will ignore this complexity. Instead we will study rather simple phenomenological neuron models, which can nevertheless reproduce the key property that allows neurons to communicate with each other - their electrical activity. It has been found that such phenomenological neuron models can reproduce neural activity to a high degree of accuracy, while still being amenable to mathematical analysis [8, 9]. The difficulty in quantifying and studying signal responses for neurons and neuron models lies in the fact that their dynamics are highly nonlinear [10]. Although, under certain assumptions, linear theories can be successfully employed to describe neural responses in experiments [11, 12], nonlinear theories are needed in order to understand and analyze the encoding and decoding of signals for different dynamic regimes of neural firing and for a wide range of stimulus parameters. It is the main aim of this thesis to provide additional tools for such a nonlinear analysis of neural data and to demonstrate under which circumstances nonlinear encoding and decoding are relevant for neural signal processing.

In particular, in chapter 2 we will introduce a *weakly nonlinear theory* for the modulation of the neural firing rate by a time-dependent signal and derive the second order response function for a neuron model analytically. We demonstrate that for cosine signals the weakly nonlinear theory is not merely a correction to the linear theory but can account for several highly pronounced nonlinear effects: excitation of harmonics, a nonlinear dependence of the mean firing rate on the signal, and a strong nonlinear response to the combination of two or more cosines. Chapter 2 is concluded by a demonstration of the presence of the nonlinear effects in a neuron model with a dynamic spiking mechanism and channel noise, hereby demonstrating the robustness and relevance of the weakly nonlinear response theory for neural firing rates. The findings presented in chapter 2 have been previously published by the author and his supervisor in [13].

Chapter 3 is devoted to the extension of the linear signal reconstruction and the linear lower bound for the mutual information. In particular, we will introduce for the special case of a static signal a *quadratic reconstruction filter* which incorporates higher-order statistics of input and output for the retrieval of information from the neural spike count. For this quadratic filter we will derive a quadratic lower bound, which indicates the information which is accessed by the quadratic filter. For a simple toy model we demonstrate how the quadratic reconstruction filter can recover information from a purely noise-coded signal - a feat at which the linear reconstruction fails. For the leaky integrate-and-fire model we demonstrate that the quadratic lower bound can exceed the linear lower bound, hereby demonstrating that a non linear filter can extract significantly more information than a linear filter does. The chapter is concluded by a demonstration of the nonlinear encoding of information in the spike count for a neuron model with a dynamic spiking mechanism and channel noise. The results from chapter 3 are currently under preparation for submission [14]. The thesis is concluded by a discussion of the results from chapter 2 and chapter 3.

In the remainder of the current chapter, we will proceed with the introduction of some

statistical tools for the analysis of neural data. We will then describe two linear theories which are commonly employed for the analysis of signal processing in neural systems: the linear response theory and the linear signal reconstruction. This introduction is concluded by a short description of the neural models which will be considered throughout this thesis.

1.1 Statistical methods for the analysis of neural data

One peculiarity about neural systems is the high variability of responses. Upon repeated presentation of the same stimulus or upon repeated injection of the same electrical current, nerve cells exhibit responses that vary from trial to trial. This variability mainly comes from three sources: i) The synapses, which are the little elements that connect two nerve cells with each other, are not always reliable. When a signal arrives at the synapse it is sometimes not transmitted [15] and sometimes a response is generated although no signal was received [16]. ii) The dynamics of the electrical activity of the nerve cells are governed by channels, which generate an ion flow across the membrane of the cell and therefore lead to an increase or a decrease of the voltage. Because there is only a finite number of channels for each cell and because a channel opens and closes in a probabilistic manner, also the voltage exhibits random fluctuations [17]. iii) The third possible source of variability arises from inputs to the nerve cell which come from other neurons of the neural network and which are not controlled by the experimenter. For many neuronal systems in cortex this background noise is the most pronounced source of variability [18], although recently there has been evidence that background activity can be also very low in the cortex of alive animals [19, 20].

Since the experimenter does not know the origin and the meaning of those inputs, they can be modeled by a stochastic noise process. In this sense *noise* only means that the purpose and origin of these inputs is not known and does not say anything about their usefulness or value for the neural system. Indeed, many studies have shown that noise can have a beneficial effect on the transmission of signals [21]. A review about the different noise sources in neural systems can be found in [22]. Because of the ubiquitous presence of noise in neural systems, neural responses cannot be quantified in a deterministic manner but have to be studied by statistical methods (cf. e.g. [23]). Some of these methods will be introduced in the following.

The electrical response emitted by each single neuron consists of stereotypic pulses that have an amplitude of about 100 mV and typically a duration of 1-2 ms. A series of such pulses or spikes is called a spike train. Since the shape of the spikes emitted by a neuron always looks alike, only the timing and number of the spikes can carry information [24]. Therefore, a spike train can be represented by a series of delta spikes

$$x(t) = \sum_{t_i} \delta(t - t_i) \ , \tag{1.1}$$

which are located at the spike times t_i. Averaging the spike train Eq. (1.1) over different independent realizations of the noise (in a biological experiment corresponding to

different trials with the same signal) we obtain the time-dependent firing rate

$$r(t) = \langle x(t) \rangle \ . \tag{1.2}$$

In this thesis, we will consider stationary conditions, i.e. conditions for which the statistics of the noise and the parameters of the neural system do not change over time. For such a setup, the time dependence in Eq. (1.2) stems solely from the applied signal. Sometimes, instead of determining the exact spike times t_i it can be more convenient to simply count the number of spikes within a time window T, which gives us the spike count

$$N(T) = \text{number of spikes within the time interval }]0, T] \ . \tag{1.3}$$

Averaging the spike count over different realizations of the noise gives us the mean spike count $\langle N(T) \rangle$. For experiments in which the signal is zero, the firing rate is time independent and denoted by r_0. For time-independent but nonzero stimuli the firing rate is also constant and will be denoted by r. For a sufficiently large observation time the constant firing rate is related to the mean spike count via a simple relation:

$$r = \frac{\langle N(T) \rangle}{T} \ . \tag{1.4}$$

Relation Eq. (1.4) implies that for time-independent signals the firing rate and the mean spike count carry the same information.

So far we only considered linear moments of the spike train and the spike count, which provide information about the average influence of the signal on the neural response. In order to quantify the variability of the neural response we also have to consider second-order moments. The variability of the spike count is usually quantified by the Fano factor [25]

$$F(T) = \frac{\langle \Delta N(T)^2 \rangle}{\langle N(T) \rangle} \ , \tag{1.5}$$

which can not only be used to quantify the variability of the count in the absence of a signal, i.e. the variability due to noise, but also to quantify the signal dependence of this variability as, for example, in [26, 27].

The second-order statistic which not only quantifies the variance of the spike train but also indicates how the values of the spike train at different times are correlated with each other is the spike-train autocorrelation function

$$R_{xx}(\tau) = \langle x(t+\tau)x(t) \rangle - \langle x(t+\tau) \rangle \langle x(t) \rangle \ . \tag{1.6}$$

Most of the time, however, the spike train autocorrelation function is hard to interpret and it is more convenient to consider the Fourier-transform of the spike train within the

time-window T,

$$\tilde{x}_T(\omega) = \int_0^T dt\; e^{i\omega t} x(t)\;. \tag{1.7}$$

Now the variance of the different Fourier components of the spike train can be quantified by the spike train power spectrum

$$S_{xx}(\omega) = \lim_{T\to\infty} \frac{\langle \tilde{x}_T(\omega)\tilde{x}_T^*(\omega)\rangle - \langle \tilde{x}_T(\omega)\rangle \langle \tilde{x}_T^*(\omega)\rangle}{T}\;, \tag{1.8}$$

where the asterisk denotes the complex conjugated. Similarly, the variance of the Fourier modes of a time-dependent signal can be quantified by the signal power spectrum

$$S_{ss}(\omega) = \lim_{T\to\infty} \frac{\langle \tilde{s}_T(\omega)\tilde{s}_T^*(\omega)\rangle - \langle \tilde{s}_T(\omega)\rangle \langle \tilde{s}_T^*(\omega)\rangle}{T}\;, \tag{1.9}$$

where $\tilde{s}$ is the Fourier-transformed signal. Note, that according to the Wiener-Khinchin theorem [28] the power spectrum can be obtained from the autocorrelation function, Eq. (1.6), via a Fourier transform. Because the Fourier transform is a linear transformation, the power spectrum contains the same information about the variability of a variable as the autocorrelation function. Only the convenience of their interpretation for a given analysis makes one preferable over the other. Further note that for stationary processes $x(t)$ and $s(t)$ and non-vanishing frequencies $f > 0$, one finds that $\langle \tilde{x}_T(\omega)\rangle = \langle \tilde{s}_T(\omega)\rangle = 0$, which in principle allows to simplify Eq. (1.8) and Eq. (1.9). However, for $f = 0$ one finds that in general $\langle \tilde{x}_T(0)\rangle \neq 0$ and $\langle \tilde{s}_T(0)\rangle \neq 0$.

1.2 Linear encoding

One of the first studies which showed that neurons respond to external sensory stimuli by a modulation of their discharge rate was the experiment conducted by Adrian on sensory neurons in the leg muscles of frogs [29]. Since then many studies have found that the firing rates of sensory neurons can be modulated by the time-course of external signals from different modalities in different animals [30, 31, 32, 33, 34]. These modulations of activity are processed by subsequent layers of neurons in the brain and ultimately lead to the generation of memories or the initiation of movement. The neurons of the different processing layers are organized in complex networks and it is the ultimate goal of neuroscience to uncover the basic principles by which these neurons process information about sensory signals. But before we can explain how networks of neurons process information we first have to understand how a signal is encoded in the response of one neuron or one layer of independent neurons. Already this seemingly simple task can pose substantial difficulties because the voltage response of neurons to incoming signals is highly nonlinear and can exhibit very complex dynamics. However, for the average rate of voltage spikes, which are generated by a neuron, in many cases one

rather simple theory proved to be very successful - the linear response theory. This theory was originally developed in quantum physics [35] but was later also applied in various other disciplines [36, 37]. The linear response theory exploits the fact that, in general, the time-dependent response, $r(t)$, of a nonlinear dynamical system which is subject to a time-dependent signal, $s(t)$, can be written as a Volterra series [38, p. 289]

$$r(t) = r_0 + \varepsilon \int_{-\infty}^{t} dt_1' \; K_1(t - t_1')s(t_1') + \varepsilon^2 \int_{-\infty}^{t}\int_{-\infty}^{t} dt_2' dt_2'' \; K_2(t - t_2', t - t_2'')s(t_2')s(t_2'') + \dots \tag{1.10}$$

where ε is a small parameter scaling the amplitude of the input signal. For weak signals ($\varepsilon \ll 1$), higher order terms in ε can be neglected and we arrive at the linear response

$$r(t) \approx r_0 + \varepsilon \int_{-\infty}^{t} dt_1' \; K_1(t - t_1')s(t_1') \; . \tag{1.11}$$

The first term in the above equation represents the output of the system in the absence of a signal. The second term in Eq. (1.11) represents the linear contribution of the input signal and its history to time-dependent fluctuations of the output at time t. The linear filter K_1 in Eq. (1.11) determines how strong past values of the signal contribute to the response and the integral essentially sums over all those contributions. For a physical system, the linear filter is a causal filter with the property that $K(t) = 0$ for $t < 0$, meaning that the response of the system can be only affected by past values of the signal but not by its future values. The notation in Eq. (1.11) is simplified if one considers the response in Fourier space

$$\begin{aligned} \tilde{r}(\omega) &= \int_{-\infty}^{\infty} dt \; e^{i\omega t} r(t) \\ &\approx 2\pi\delta(\omega) r_0 + \varepsilon \chi_1(\omega)\tilde{s}(\omega) \; , \end{aligned} \tag{1.12}$$

where $\tilde{s}$ is the Fourier-transformed signal and χ_1 is the Fourier transform of the linear filter K_1

$$\chi_1(\omega) = \int_{-\infty}^{\infty} dt \; e^{i\omega t} K_1(t) \; . \tag{1.13}$$

Apart from the delta peak at $\omega = 0$, the Fourier-transformed linear response, Eq. (1.12), can be represented by a simple product of two functions without any integration. For this reason, it is often convenient to study neural responses in the Fourier space. The specific shape of $\chi_1(\omega)$ depends on the properties of the dynamical system under study and can be either measured numerically for physical or biological systems or computed analytically for mathematical models.

In the context of neuroscience linear response theory was successfully applied to study properties of signal transmission in real neurons [11, 12, 39] and for analytic calculations of the response of neural models [40, 41, 42]. Some questions that could be tackled by the use of the linear response theory are, for example, the transmission of fast signals [43, 44] and the stability of recurrent neural networks [45]. Despite its success, however, linear response theory cannot describe the full dynamic range of nonlinear systems and, consequently, extensions of the linear response theory have been studied e.g. in nonlinear optics [46, 47, 48] and magnetic particle imaging [49]. For neural systems the nonlinear response was measured experimentally [50, 51, 52, 53], was employed to study the network stability of a recurrent network [54] and even calculated analytically for a simple Poisson neuron model [55]. However, a complete theory for the second-order response for a neuron model with explicit voltage dynamics is still missing and it is still an open question whether the second-order response bears any significance for biological neurons. It is the aim of chapter 2 to fill this gap, to derive a complete second order response theory for the leaky integrate-and-fire model, to identify the second-order effects of a signal on the neural output and to demonstrate under which conditions the second-order effects are expected to be most pronounced in biological neurons.

1.3 Linear decoding

In the previous section we talked about the linear encoding of signals, i.e. we talked about the prediction of the statistics of the output for a given input signal. In this section we will consider the reverse approach and discuss how it is possible to reconstruct the input signal from an observation of the neural output. One possible approach is the linear signal reconstruction, which is discussed in detail in [38] and which we will shortly introduce in the following.

Linear signal reconstruction

We observe a spike train and we would like to reconstruct the signal that led to the generation of this spike train. The difficulty in the reconstruction lies in the nonlinear nature of the spike generation mechanism in neural systems and the ubiquitous influence of neural noise, which we discussed above. Consequently, two questions arise: i) How much information about the input signal is contained int the neural output? ii) How can we extract this information and how can we use it for the reconstruction of the signal?

First we attempt to write the reconstructed input signal, s_{rec}, as a functional of the neural response. We will try to "guess" a mathematical operation that extracts information about the input signal from the neural response and will in the end compare the estimated signal with the real signal that was presented to the neural system. A rather simple yet very effective choice is a linear filter operation [56], for which the reconstructed

signal is given by

$$s_{rec}(t) = \int_{-\infty}^{\infty} d\tau \; h(t-\tau)x(\tau) \; . \tag{1.14}$$

The linear kernel $h(t)$ in the above equation is determined by an optimization routine which maximizes the quality of the reconstruction. Therefore, Eq. (1.14) is sometimes referred to as the optimal linear reconstruction. Note, that Eq. (1.14) is essentially a Volterra expansion of $s_{rec}(t)$ with respect to $x(t)$ which is truncated after the first-order term and is in this sense similar to the linear response in Eq. (1.11). The difference between the two Volterra expansions, however, is that in contrast to Eq. (1.11) the linear kernel in Eq. (1.14) is in general not causal, because the neural output at $t' > t$ is in general contributing to the reconstruction of the signal at time t. In order to compute the optimal linear kernel $h(t)$, we first Fourier transform Eq. (1.14) into

$$\tilde{s}_{rec}(\omega) = \tilde{h}(\omega)\tilde{x}(\omega) \; , \tag{1.15}$$

such that we get rid of the integral in Eq. (1.14). Now, the optimal linear filter is determined by minimizing the variance of the reconstruction error $\tilde{\eta}(\omega) = \tilde{s}(\omega) - \tilde{s}_{rec}(\omega)$. The filter that is obtained in this way is called the optimal Wiener filter and reads [38]

$$\tilde{h}(\omega) = \frac{S^*_{xs}(\omega)}{S_{xx}(\omega)} \tag{1.16}$$

where

$$S_{xs}(\omega) = \lim_{T\to\infty} \frac{\langle \tilde{x}_T(\omega)\tilde{s}^*_T(\omega)\rangle - \langle \tilde{x}_T(\omega)\rangle \langle \tilde{s}^*_T(\omega)\rangle}{T} \tag{1.17}$$

is the cross spectrum between the spike train and the signal. By inserting $\tilde{h}$ from Eq. (1.16) into Eq. (1.15) it is possible to use Eq. (1.15) for the reconstruction of the input signal from an observation of the output $\tilde{x}$.

The quality of the resulting reconstruction can be assessed via the squared linear correlation function between the real signal $\tilde{s}(\omega)$ and the reconstructed signal $\tilde{s}_{rec}(\omega)$

$$C_{s,s_{rec}}(\omega) = \frac{|\, \langle \tilde{s}(\omega)\tilde{s}^*_{rec}(\omega)\rangle - \langle \tilde{s}(\omega)\rangle \langle \tilde{s}^*_{rec}(\omega)\rangle \,|^2}{(\, \langle \tilde{s}(\omega)\tilde{s}^*(\omega)\rangle - \langle \tilde{s}(\omega)\rangle \langle \tilde{s}^*(\omega)\rangle \,)(\, \langle \tilde{s}_{rec}(\omega)\tilde{s}^*_{rec}(\omega)\rangle - \langle \tilde{s}_{rec}(\omega)\rangle \langle \tilde{s}^*_{rec}(\omega)\rangle \,)} \; . \tag{1.18}$$

Using Eq. (1.15) and Eq. (1.16) it is possible to show that Eq. (1.18) can also be written as

$$C_{s,s_{rec}}(\omega) = \frac{|S_{xs}(\omega)|^2}{S_{xx}(\omega)S_{ss}(\omega)} \; . \tag{1.19}$$

The expression on the right hand side of Eq. (1.19) is the coherence function between

input and output

$$C_{xs}(\omega) = \frac{|S_{xs}(\omega)|^2}{S_{xx}(\omega)S_{ss}(\omega)}. \tag{1.20}$$

From the right hand side of Eq. (1.19), we can see that in order to determine whether a signal can be reconstructed linearly from the output or not, we don't have to actually perform the reconstruction. Instead it is sufficient to measure the coherence function, which requires only the knowledge of the second-order statistics of input and output, in order to assess the potential quality of the reconstruction. The coherence function quantifies linear correlations of the frequency components of the signal with the respective frequency components of the spike train. It attains values between zero and one, where values of the coherence which are close to one indicate that the linear reconstruction successfully extracts information about the signal from the neural spike train. For time-independent stimuli which are encoded in the spike count Eq. (1.14) reduces to

$$s_{rec} = h \cdot N \tag{1.21}$$

with

$$h = \frac{\langle Ns \rangle - \langle N \rangle \langle s \rangle}{\langle N^2 \rangle - \langle N \rangle^2} \ . \tag{1.22}$$

One of the first successful demonstrations of a reconstruction of a visual stimulus from the spike trains of a neuron in the visual system of the fly was performed by Bialek et al. [12]. Since then it has been demonstrated in many studies that the linear signal reconstruction can successfully extract information about the signal from the neural output. However, there is one problem with the linear signal reconstruction and the coherence function (or the linear correlation coefficient for static signals): It is not clear how to interpret low values of these functions. On the one hand, low values could mean that there is only little information about the signal in the spike train and that most of the spikes are caused by the internal noise, on the other hand low values could mean that the information in the neural output is not accessible to a linear filter but could potentially be retrieved by other (nonlinear) methods. Indeed, it has been found that the linear filter works well for neural systems with weak input signals but fails for strong input signals, where the signal encoding is expected to be nonlinear. Consequently, the following question arises: How can we quantify the total information that is contained in the output without making assumptions about the neural code or about possible decoding mechanisms? The answer to this question is provided by information theory, which we will introduce in the following section.

Information theory

In order to be able to quantify the information that a neural output contains about an input signal, we first need a definition of what information actually is. A formal

definition of information, that can be applied to statistical codes (such as the neural code) was provided by Shannon in 1948 [57], the main concepts of which will be briefly introduced in the following.

Consider a discrete output variable N. Given that we know the statistical properties of N, we can assign to this variable a quantity called the Shannon entropy,

$$H_N = -\sum_i P(N_i) \log_2 \left(P(N_i)\right) \ , \tag{1.23}$$

where $P(N_i)$ is the probability that the variable N will attain the state i and where the sum runs over all possible states. The entropy Eq. (1.23) gives the total variability of the output measured in bits and indicates the minimum average number of binary questions ("yes" / "no" questions) that are needed to determine the exact state of the variable N [58]. Alternatively, we can interpret Eq. (1.23) as an average of the expression $\log_2(1/P_i)$, which in turn can be interpreted as a measure of the surprise or uncertainty associated with the state i. The less likely a state is, the more surprising it is to observe that state upon an observation of the system and the more uncertainty we have about N before we observe its state. Therefore, Eq. (1.23) can be interpreted as the average surprise or the average uncertainty associated with the random variable N.

If N is the output of a system that is subject to an input signal s and if we know the value of s for a given realization of the output N, then in general our surprise about N will be reduced. The average surprise associated with the random variable N given that we know the input signal s, i.e. the entropy of N conditioned on s, is defined as follows:

$$H_{N|s} = -\sum_j P(s_j) \sum_i P(N_i|s_j) \log_2 \left(P(N_i|s_j)\right) \ , \tag{1.24}$$

where $P(N_i|s_j)$ is the conditional probability that N is in state i given that s is in state j. If the state of the signal is known, then the only source for the variability of the output is the intrinsic noise of the system. Therefore, Eq. (1.24) is also referred to as the noise entropy. Analogously to the entropy in Eq. (1.23), Eq. (1.24) can be interpreted as the minimum average number of binary questions that are needed to determine the exact state of N given that we know the state of the signal. Because the knowledge of the signal already provides some information about the output, the number of question needed to determine the state of N will be reduced as compared to Eq. (1.23).

The key property of information theory, the mutual information between the input s and output N, is defined by the difference of the entropies in Eq. (1.23) and Eq. (1.24):

$$MI = H_N - H_{N|s} \ . \tag{1.25}$$

From the above equation we find that in the framework of information theory, information is defined as the mean reduction of uncertainty about the output variable N upon observation of the input variable s. Using the Bayes' theorem,

$$P(N_i|s_j)P(s_j) = P(N_i, s_j) = P(s_j|N_i)P(N_i) \ , \tag{1.26}$$

we can rewrite Eq. (1.25) into

$$MI = \sum_j \sum_i P(N_i, s_j) \log_2 \left(\frac{P(N_i, s_j)}{P(N_i)P(s_j)} \right) , \tag{1.27}$$

which shows that the mutual information is a symmetric quantity with respect to signal and output. In other words, the reduction of uncertainty that a signal is providing about the neural output (signal encoding) equals the reduction of uncertainty that a neural output is providing about the signal (signal decoding). In this sense, signal encoding and signal decoding are simply two sides of the same coin. Furthermore, looking at Eq. (1.27) we can interpret the mutual information simply as a nonlinear measure of the statistical dependencies between the signal s and the output N which are captured by the joint probability $P(N_i, s_j)$. For statistically independent variables one finds that $P(N_i, s_j) = P(N_i)P(s_j)$ and the mutual information in Eq. (1.27) is zero.

So far we have considered only two variables, e.g. a static signal and the neural spike count, in which case each of the single variables can be quantified by univariate probability distributions. For time-dependent signals and outputs, it is necessary to perform a time-discretization, such that the signal and output are represented by random vectors, $\vec{s}$ and $\vec{x}$, and different elements of the vectors correspond to the values of signal and output in different time-bins. For the time-dependent case, the input and output are quantified by multivariate distributions and the mutual information is given by

$$MI = \sum_i \sum_j P(\vec{x}_i, \vec{s}_j) \log_2 \left(\frac{P(\vec{x}_i, \vec{s}_j)}{P(\vec{x}_i)P(\vec{s}_j)} \right) , \tag{1.28}$$

where i and j now correspond to different states of the signal and output vectors. For continuously distributed variables, the entropies are referred to as *differential entropies* [23] and are defined as

$$H_x = \int_{-\infty}^{\infty} dx\, p(x) \log_2 p(x) , \quad \left\langle H_{x|s} \right\rangle_s = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ds\, dx\, p(x, s) \log_2 p(x|s) . \tag{1.29}$$

The mutual information is then given by

$$MI = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ds\, dx\, p(s, x) \log_2 \left(\frac{p(x, s)}{p(x)p(s)} \right) . \tag{1.30}$$

For time-dependent signals and neural spike trains, the mutual information has been measured in some experimental studies [59, 60], although the estimation of the probabilities of the states is not straightforward [61] and can only be performed for neural systems for which a lot of data can be recorded [62]. Nevertheless, the information theory by Shannon provides a rigorous definition of information and allows to quantify the full information that the neural output caries about the signal. In particular, the mutual

information does not rely on assumptions about the underlying "neural code" but simply accounts for all (potentially nonlinear) statistical dependencies between signal and output. In this sense, the mutual information is a nonlinear measure of signal transmission and a comparison of the mutual information with linear information measures can reveal regimes in which signal encoding is nonlinear.

Linear lower bound for the mutual information

In the previous section we introduced the mutual information, which allows to determine the total information that is contained in the neural output about an input signal. The mutual information can be interpreted as the maximal amount of information about a stimulus which can potentially be extracted from the output either by a mathematical operation or by a biological readout mechanism in the brain. The knowledge of the maximally available information is a prerequisite for the evaluation of biological readout mechanisms or for candidate neural codes which may or may not actually use this information.

The minimal rate at which information can be recovered by the linear signal reconstruction Eq. (1.14) is given by the lower bound for the mutual information rate [12, 38]

$$R_{lb}^{lin} = -\frac{1}{2\pi}\int_0^\infty d\omega \ \log_2\left(1 - C(\omega)\right) \ . \tag{1.31}$$

The linear lower bound can be formally shown to be always smaller than or equal to the exact mutual information rate [63]. For a static signal, Eq. (1.31) reduces to

$$R_{lb}^{lin} = -\frac{1}{2T}\log_2\left(1 - \rho_{s,N}^2\right) \ , \tag{1.32}$$

where $\rho_{s,N}$ is the Pearson's correlation coefficient between the signal and the output. For different neuronal systems it has been shown that the linear lower bound comes very close to the exact mutual information, indicating that a linear reconstruction can extract almost all of the information that is present in the output [62]. Indeed, the close match between the linear lower bound and the exact mutual information has been so compelling, that the linear lower bound has been often used as an estimate for the exact mutual information. In recent years, however there is more and more evidence, that the linear reconstruction performs sub optimally and that the exact mutual information can be much larger than the linear lower bound [64, 65, 66]. If the linear reconstruction fails to extract all of the information that is in the neural output, the question arises by which mechanisms this information can be accessed. It is the aim of chapter 3 of this thesis to provide a nonlinear extension to the linear theory and to demonstrate the significance of the nonlinear reconstruction for neural systems.

1.4 Neuron models

The goal of this thesis is to contribute to a better understanding of the basic principles of nonlinear signal processing in neural systems. Therefore, it is not appropriate to study complex neural models for which we have no chance to uncover any basic principles. Instead we follow for the selection of the complexity of a neural model the following simple rule: The model should be "simple enough to be useful but complex enough to be meaningful" [9]. The leaky integrate-and-fire (LIF) model satisfies this rule perfectly. It ignores the complex morphology and the detailed biology of nerve cells and is easy enough to be amenable to a mathematical analysis. On the other hand it is complex enough to reproduce certain features of neural responses quite well [8, 9]. In particular, the LIF model is able to reproduce certain features of nonlinear signal transmission of real neurons [67]. The effects which we will first discuss for the LIF model will also be demonstrated for the sodium and potassium (NaK) neuron model. In contrast to the LIF model, the NaK model exhibits implicit dynamics for the generation of spikes and is endowed with discrete channel noise. The demonstration of nonlinear signal encoding and decoding for the NaK model will demonstrate that our analytical and theoretical results do not hinge on the particularities of one single model but are of a more general nature.

Leaky integrate-and-fire (LIF) model

The historical origin of the LIF neuron model dates as far back as 1907 [68]. Nevertheless, although more than a century has passed since then, the model has not lost its appeal until today (see [69] for a historical overview). The LIF model can reproduce some aspects of the nonlinear behavior of real nerve cells. At the same time the model is still amenable to an extensive mathematical analysis. For example, the LIF model has been successfully employed for the mathematical study of signal transmission in single neurons (e.g. in [40, 41]) and for the mathematical description of neural networks (e.g. in [45, 54, 70, 71]). The time-dependent voltage dynamics of the LIF model is given by

$$\tau_m \dot{v} = -v + \mu + \epsilon s(t) + \sqrt{2D}\xi(t) \;, \tag{1.33}$$

which is complemented by a fire and reset rule: Whenever the voltage hits the threshold v_T a spike is registered and the voltage is reset to the reset value v_R. For the duration of the refractory time τ_r the voltage is fixed to the reset value and only after the refractory time has passed, the voltage again follows Eq. (1.33). In this thesis we will consider a dimensionless variant of Eq. (1.33) with $v_T = 1$ and $v_R = 0$. An example of a signal, a voltage trajectory, spike trains and the time-dependent firing rate is shown in Fig. 1.1. The noise $\xi(t)$ in Eq. (1.33) is a Gaussian white noise with zero mean that is uncorrelated in time. The noise strength D scales the strength of the noise process and the parameter μ corresponds to the mean current that is applied to the neuron. For the mathematical analysis and for most of the simulations in this thesis, we will set the membrane time-constant, τ_m, in Eq. (1.33) to one. Consequently, time in our analysis is given in units

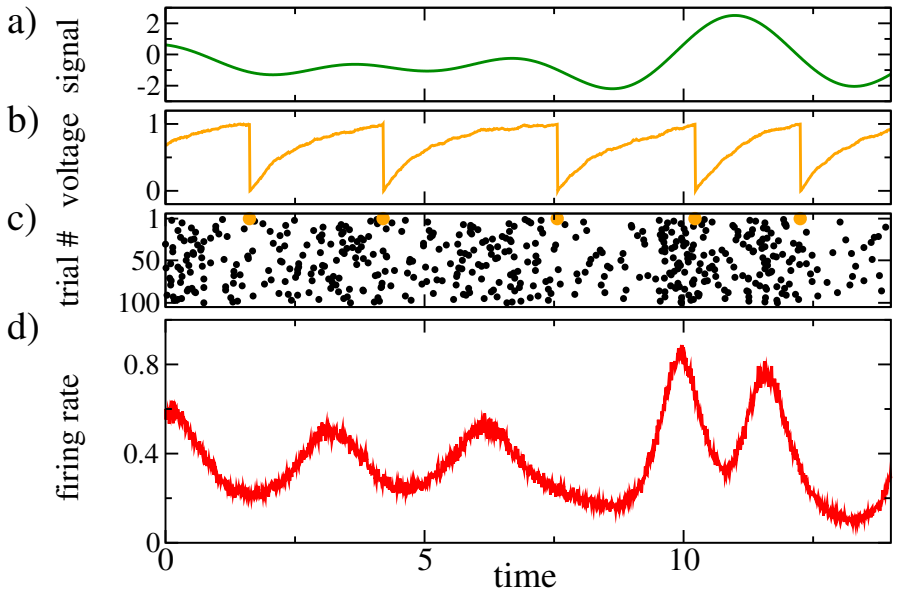

Figure 1.1: Input signal, voltage, spike trains and firing rate for the LIF model. (a) Gaussian signal with a constant non-vanishing power spectrum for frequencies up to $f_c = 0.3$ and vanishing power for $f > f_c$. (b) voltage; (c) spike trains for different trials (small black dots); The spike train corresponding to the voltage in b) is shown in the first row (large orange dots). (d) Time-dependent firing rate. The firing rate is obtained by trial averaging over 100000 spike trains, the first 100 of which are shown in (c). Parameters are $\varepsilon = 0.05$, $\mu = 1.1$, $D = 0.001$.

of τ_m and the firing rate becomes a dimensionless quantity. For $\mu > v_T$ the voltage in Eq. (1.33) always runs towards the threshold, even for $D = 0$. Therefore, this regime is called the mean-driven regime. For $\mu < v_T$ the voltage of the neuron cannot cross the threshold if the intrinsic noise is switched off ($D = 0$). Only for $D > 0$ the neuron can cross the threshold due to fluctuations caused by the noise. Consequently, the regime with $\mu < v_T$ is called the fluctuation-driven regime. In this thesis we will show that the mean-driven and the fluctuation-driven regimes have very different properties with respect to nonlinear signal processing.

Although Eq. (1.33) is usually used to implement the numerical simulation of the LIF model, for a mathematical analysis of the statistics of the neural response it is often more convenient to consider a differential equation for the time-dependent density of the voltage $p(v,t)$. Such an equation is called Fokker-Planck equation [28, 37] and for the LIF model it is given by [72]

$$\partial_t p(v,t) = \hat{L}_0 p(v,t) - \epsilon s(t) \partial_v p(v,t) + r(t-\tau_r)\delta(v-v_R) \,, \tag{1.34}$$

where $\hat{L}_0 = \partial_v (v - \mu + D\partial_v)$. In addition to Eq. (1.34) the density $p(v,t)$ has to obey boundary conditions and a normalization condition. First, an absorbing boundary con-

dition at the threshold leads to

$$p(v_T, t) = 0 \tag{1.35}$$

and for the derivative at the threshold the absorbing boundary conditions implies

$$-Dp'(v_T, t) = r(t) \ , \tag{1.36}$$

where the prime denotes a partial derivative with respect to v. Furthermore, the density is continuous everywhere, which in particular at v_R leads to

$$\lim_{\Delta v \to 0} \left(p(v_R + \Delta v, t) - p(v_R - \Delta v, t) \right) = 0 \ . \tag{1.37}$$

The density should also be integrable implying

$$\lim_{v \to -\infty} p(v, t) = 0 \ . \tag{1.38}$$

Finally, the density is normalized to one

$$\int_{-\infty}^{v_T} dv \ p(v, t) + \int_{t-\tau_r}^{t} dt' \ r(t') = 1 \ , \tag{1.39}$$

where the last term in the above equation is taking into account realizations of the model that are in the refractory state at time t. As we will discuss later in detail, the Fokker-Planck formalism allows to compute the unperturbed firing rate [73, 74]

$$r_0 = \left[\tau_r + \sqrt{\pi} \int_{(\mu - v_R)/\sqrt{2D}}^{(\mu - v_T)/\sqrt{2D}} dx \ e^{x^2} \operatorname{erfc}(x) \right]^{-1} \ , \tag{1.40}$$

which is the firing rate of the neuron in the absence of any signals. The power spectrum can be also be computed from the Fokker-Planck formalism and reads [75]

$$S_{xx}(\omega) = r_0 \frac{\left| \mathcal{D}_{i\omega}\left(\frac{\mu - v_T}{\sqrt{D}}\right) \right|^2 - e^{2\Delta} \left| \mathcal{D}_{i\omega}\left(\frac{\mu - v_R}{\sqrt{D}}\right) \right|^2}{\left| \mathcal{D}_{i\omega}\left(\frac{\mu - v_T}{\sqrt{D}}\right) - e^{\Delta} e^{i\omega\tau_r} \mathcal{D}_{i\omega}\left(\frac{\mu - v_R}{\sqrt{D}}\right) \right|^2} \tag{1.41}$$

where $\Delta = (v_R^2 - v_T^2 + 2\mu(v_T - v_R))/(4D)$ and where the function $\mathcal{D}$ is the parabolic cylinder function [76]. The analytical result for the linear response function for the LIF model [41, 42] is given by

$$\chi_1(\omega) = \frac{r_0 i\omega/\sqrt{D}}{i\omega - 1} \frac{\mathcal{D}_{i\omega-1}(\frac{\mu - v_T}{\sqrt{D}}) - e^{\Delta} \mathcal{D}_{i\omega-1}(\frac{\mu - v_R}{\sqrt{D}})}{\mathcal{D}_{i\omega}(\frac{\mu - v_T}{\sqrt{D}}) - e^{\Delta} e^{i\omega\tau_r} \mathcal{D}_{i\omega}(\frac{\mu - v_R}{\sqrt{D}})} \ . \tag{1.42}$$

The explicit expression for the unperturbed rate, the power spectrum and the linear

response function are very helpful for the mathematical analysis of signal transmission in single neurons but also for the study of the dynamics of neural networks. The expressions in Eqs. (1.40-1.42) can also be evaluated numerically by different programming languages, which is usually faster and less error prone than measuring these quantities from numerical simulations of Eq. (1.33). In chapter 2 we will employ the Fokker-Planck formalism for the derivation of the nonlinear response of the LIF model and will employ Eq. (1.40) and Eq. (1.42) for the subsequent quantitative analysis of the nonlinear signal transmission for the LIF model. The expression for the power spectrum Eq. (1.41) will be used in chapter 3.

Izhikevich's sodium-potassium (NaK) model with ion channel noise

The NaK model was introduced by Izhikevich in [10] and is similar to the two-dimensional Morris-Lecar model [77]. Here, we consider a type of the NaK model which is subject to ion channel noise [78]. This will allow us to test whether nonlinear signal processing, which we will demonstrate for the LIF model, can also be observed in a model with a more biologically plausible type of noise. The voltage dynamics of the NaK model are given by

$$C\dot{v} = I_0 + \varepsilon s(t) - I_L(v) - I_{Na}(v) - I_K(v, n(t)) \,, \tag{1.43}$$

where I_0 is a constant input current, I_L is a passive leak current, I_{Na} is a deterministic "persistent sodium" current and I_K is a potassium current gated by the number of open potassium channels $0 \leq n(t) \leq n_{tot}$ where n_{tot} is the total number of channels. The number of open channels $n(t)$ comprises a time-dependent stochastic process with voltage dependent per capita transition rates $\rho_o(v)$ for channel opening and $\rho_c(v)$ for channel closing. In our simulations, we consider a potassium current with a high threshold and the following parameters: Passive leak current $I_L = g_L(v - E_L)$ with $g_L = 8\text{mS/cm}^2$ and $E_L = -80\text{mV}$, "persistent sodium" current $I_{Na} = g_{Na} m_\infty(v)(v - E_{Na})$ with $g_{Na} = 20\text{mS/cm}^2$, $E_{Na} = 60\text{mV}$ and voltage dependent activation $m_\infty(v) = 1/\{1 + \exp[(-20\text{mV} - v)/15\text{mV}]\}$; potassium current $I_K = (n(t)/N) g_k(v - E_K)$ with $g_K = 10\text{mS/cm}^2$, $E_K = -90\text{mV}$. Membrane capacitance is $C = 1\mu\text{F/cm}^2$. The per capita transition rate for channel opening is $\rho_o(v) = 1/\{1 + \exp[(-25\text{mV} - v)/5\text{mV}]\}$ and for closing is $\rho_c(v) = 1 - \rho_o(v)$. The total number of potassium channels is $n_{tot} = 100$. Whenever the voltage in Eq. (1.43) exceeds 0mV a spike is registered. An example of an input signal, a voltage trajectory, spike trains and the resulting time-dependent firing rate are shown in Fig. 1.2.

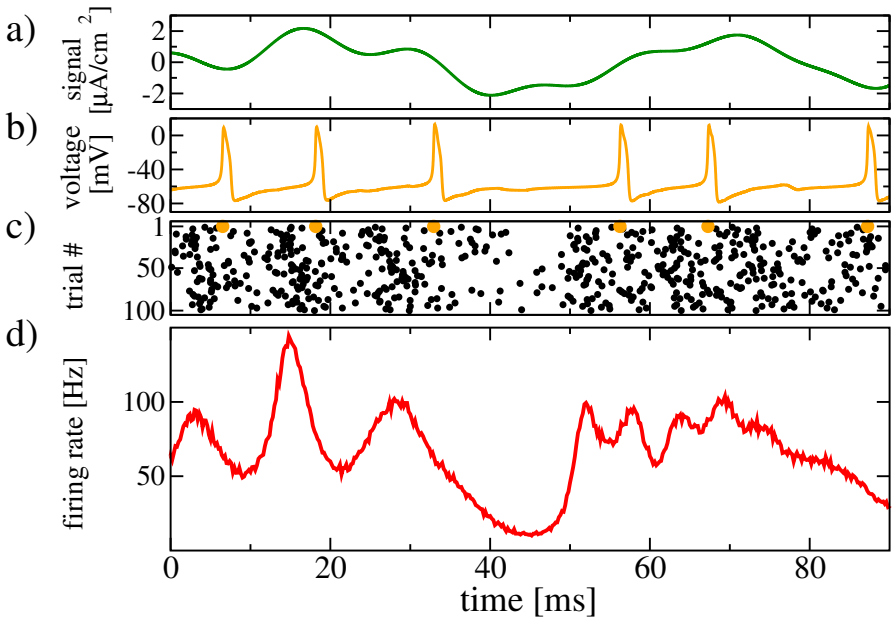

Figure 1.2: Input signal, voltage, spike trains and firing rate for the NaK model. (a) Gaussian signal with a constant non-vanishing power spectrum for frequencies up to $f_c = 80$ Hz and vanishing power for $f > f_c$. (b) voltage; (c) spike trains for different trials (small black dots); The spike train corresponding to the voltage in b) is shown in the first row (large orange dots). (d) Time-dependent firing rate. The firing rate is obtained by trial averaging over 100000 spike trains, the first 100 of which are shown in (c). Parameters are $\varepsilon = 0.3$, $I_0 = 5\mu\text{A/cm}^2$.

2 Nonlinear encoding of time-dependent signals in the neural firing rate

Abstract | In this chapter we analytically derive the second-order response function for the LIF model. This function allows to predict the time-dependent firing rate in response to general time-dependent signals, which is not only an interesting single-neuron statistics that has been measured in experiments, but can also be regarded as the instantaneous population rate of a group of uncoupled neurons subject to the same external drive. For cosine signals, The weakly nonlinear theory successfully describes the excitation of higher harmonics, a frequency-dependent change of the time-averaged firing rate and a strong nonlinear interaction of two signals with different frequencies. The chapter is concluded by a discussion of a numerical routine for the estimation of the second order response function from neural data and by an application of this routine to the NaK neuron model.

For sufficiently weak signals ($\varepsilon \ll 1$) the time-dependent firing rate of a neuron, $r(t)$, can be represented by the Volterra expansion in Eq. (1.10). Sometimes it is more convenient to represent the firing rate in Fourier space, in which case Eq. (1.10) turns into

$$\tilde{r}(\omega) = 2\pi\delta(\omega)r_0 + \epsilon\chi_1(\omega)\tilde{s}(\omega) + \epsilon^2\frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega'\chi_2(\omega-\omega',\omega')\tilde{s}(\omega')\tilde{s}(\omega-\omega') + \dots\ , \quad (2.1)$$

where

$$\chi_1(\omega) = \int_{-\infty}^{\infty} dt\ e^{i\omega t}K_1(t) \quad (2.2)$$

and

$$\chi_2(\omega_1,\omega_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dt_1 dt_2\; e^{i\omega_1 t_1} e^{i\omega_2 t_2} K_2(t_1,t_2)\;. \tag{2.3}$$

The advantage of Eq. (2.1) over Eq. (1.10) lies in the often simpler mathematical representation of the functions χ_1 and χ_2 as opposed to the kernels K_1 and K_2 and in the reduced number of integrations which are involved in the computation of the response. Once $\tilde{r}(\omega)$ is evaluated, it is always possible to perform a numerical Fourier back transformation which is available in the standard libraries of many programming languages. In the following we will refer to the theory in Eq. (2.1) with terms up to second order in ε as the weakly nonlinear theory. Leaving out the terms which are proportional to ε^2 and only keeping zero-order and first-order terms in Eq. (2.1), we can obtain the well-known linear response theory [35, 36, 37].

The knowledge of the functions r_0, χ_1, and χ_2 allows to compute the second-order response of a neuron or neuron model for arbitrary time-dependent signals. For the leaky integrate-and-fire model these functions can be determined by solving the Fokker-Planck equation Eq. (1.34), which will be demonstrated in the following. Hereby, we will first reproduce known results for the first-order response function [41, 42] and will then provide an explicit expression for the full second-order response function $\chi_2(\omega_1,\omega_2)$, which has also been published by the author and his supervisor in [13]. Note, that the weakly nonlinear regime which has been studied before in [54] only involved the response to a single cosine and did not describe the full second-order response which will be presented here. After the mathematical derivation we will briefly discuss the symmetries and some other properties of the function $\chi_2(\omega_1,\omega_2)$ and will then proceed to study the weakly nonlinear response of the LIF model. For the special case of cosine signals, we will demonstrate the following nonlinear effects: i) a frequency-dependent change of the time-averaged firing rate reminiscent of frequency locking in deterministic oscillators; ii) higher harmonics in the rate that may exceed the linear response; iii) a strong nonlinear response to two cosines even when the response to a single cosine is linear. We will proceed with an investigation of the influence of the signal amplitude and the strength of the background noise on the emergence of the second-order effects for the LIF model. Finally, we will conclude this chapter by demonstrating that the second-order effects which we observe for the LIF model are also observed for the NaK model with channel noise hereby demonstrating the robustness of the weakly nonlinear effects.

2.1 Derivation of the second-order response function for the LIF model

Now, we want to find an ansatz for the time-dependent firing rate, $r(t)$, and the density $p(v,t)$, which will allow us to compute the second-order response for the LIF model. In

other words, we want to compute the functions $K_1(t)$ and $K_2(t_1, t_2)$ from Eq. (1.10) or the corresponding Fourier transforms $\chi_1(\omega)$ and $\chi_2(\omega_1, \omega_2)$. In order to obtain a suitable ansatz we consider the signal

$$s(t) = \alpha \cos(\omega_1 t) + \beta \cos(\omega_2 t + \varphi) \,, \tag{2.4}$$

where α and β are constants which scale the amplitudes of the respective cosines and φ is the relative phase between the two cosines. Inserting Eq. (2.4) into the Volterra series for the time-dependent firing rate, Eq. (1.10), and exploiting the causality of the kernels leads to

$$\begin{aligned} r(t) =& r_0 + \frac{\epsilon}{2} \int_{-\infty}^{\infty} dt_1' \; K_1(t_1') \left[\alpha e^{-i\omega_1 t} e^{i\omega_1 t_1'} + \beta e^{-i\omega_2 t} e^{i\omega_2 t_1'} e^{-i\varphi} + c.c. \right] \\ &+ \frac{\epsilon^2}{4} \int_{-\infty}^{\infty} dt_2' \int_{-\infty}^{\infty} dt_2'' \; K_2(t_2', t_2'') \left[\alpha e^{-i\omega_1 t} e^{i\omega_1 t_2'} + \beta e^{-i\omega_2 t} e^{i\omega_2 t_2'} e^{-i\varphi} + c.c. \right] \\ &\times \left[\alpha e^{-i\omega_1 t} e^{i\omega_1 t_2''} + \beta e^{-i\omega_2 t} e^{i\omega_2 t_2''} e^{-i\varphi} + c.c. \right] \,, \end{aligned} \tag{2.5}$$

where by *c.c.* we denote the complex conjugated. Employing the definitions of the Fourier transforms of the kernels K_1 and K_2 from Eq. (2.2) and Eq. (2.3), the firing rate in Eq. (2.5) can be simplified into

$$\begin{aligned} r(t) &= r_0 + \frac{\varepsilon}{2} \left(\alpha \chi_1(\omega_1) e^{-i\omega_1 t} + \beta \chi_1(\omega_2) e^{-i\omega_2 t} e^{-i\varphi} + c.c. \right) \\ &+ \frac{\varepsilon^2}{2} \left(\alpha^2 \chi_2(\omega_1, -\omega_1) + \beta^2 \chi_2(\omega_2, -\omega_2) \right) \\ &+ \frac{\varepsilon^2}{4} \left(\alpha^2 \chi_2(\omega_1, \omega_1) e^{-i2\omega_1 t} + \beta^2 \chi_2(\omega_2, \omega_2) e^{-i2\omega_2 t} e^{-i2\varphi} + c.c. \right) \\ &+ \frac{\varepsilon^2 \alpha\beta}{2} \left(\chi_2(\omega_1, \omega_2) e^{-i(\omega_1+\omega_2)t} e^{-i\varphi} + \chi_2(\omega_1, -\omega_2) e^{-i(\omega_1-\omega_2)t} e^{i\varphi} + c.c. \right) \,. \end{aligned} \tag{2.6}$$

The above equation will now be the ansatz for $r(t)$ which we will employ in Eq. (1.34) for the perturbation calculation of the second-order response function. Note, that we choose the signal, Eq. (2.4), such that the resulting ansatz allows to compute the complete function $\chi_2(\omega_1, \omega_2)$. If we chose a signal consisting of only one cosine, we would only obtain $\chi_2(\omega, \omega)$, i.e. we would obtain the second-order response function only for certain frequency arguments but not the complete function. In turn, choosing a more complex signal would not provide any additional information about the second-order response and would only complicate the calculation. Also note, that once we have calculated the full function $\chi_2(\omega_1, \omega_2)$ we can determine the second-order response to any arbitrary signal by virtue of Eq. (2.1). The ansatz for the density $p(v, t)$ is chosen analogously to

Eq. (2.6) and reads

$$\begin{aligned} p(v,t) = \; & p_0(v) + \frac{\varepsilon}{2}\left(\alpha p_1(v,\omega_1)e^{-i\omega_1 t} + \beta p_1(v,\omega_2)e^{-i\omega_2 t}e^{-i\varphi} + c.c.\right) \\ & + \frac{\varepsilon^2}{2}\left(\alpha^2 p_2(v,\omega_1,-\omega_1) + \beta^2 p_2(v,\omega_2,-\omega_2)\right) \\ & + \frac{\varepsilon^2}{4}\left(\alpha^2 p_2(v,\omega_1,\omega_1)e^{-i2\omega_1 t} + \beta^2 p_2(v,\omega_2,\omega_2)e^{-i2\omega_2 t}e^{-i2\varphi} + c.c.\right) \\ & + \frac{\varepsilon^2\alpha\beta}{2}\left(p_2(v,\omega_1,\omega_2)e^{-i(\omega_1+\omega_2)t}e^{-i\varphi} + c.c.\right) \\ & + \frac{\varepsilon^2\alpha\beta}{2}\left(p_2(v,\omega_1,-\omega_2)e^{-i(\omega_1-\omega_2)t}e^{i\varphi} + c.c.\right) \; . \end{aligned} \tag{2.7}$$

Now we insert Eq. (2.6) and Eq. (2.7) into the Fokker-Planck equation, Eq. (1.34), and obtain, after sorting the resulting terms according to their time dependence, the following hierarchy of equations:

$$0 = \hat{L}_0 p_0(v) + r_0\delta(v - v_R), \tag{2.8}$$

$$0 = (\hat{L}_0 + i\omega)p_1(v,\omega) + \chi_1(\omega)e^{i\omega\tau_r}\delta(v - v_R) - \partial_v p_0(v) \; , \tag{2.9}$$

$$\begin{aligned} 0 = \; & (\hat{L}_0 + i(\omega_1{+}\omega_2))p_2(v,\omega_1,\omega_2) + \chi_2(\omega_1,\omega_2)e^{i(\omega_1+\omega_2)\tau_r}\delta(v - v_R) \\ & - \frac{1}{2}\big(\partial_v p_1(v,\omega_1) + \partial_v p_1(v,\omega_2)\big) \; . \end{aligned} \tag{2.10}$$

The boundary conditions for the densities $p_{0,1,2}$ are determined by inserting Eq. (2.6) and Eq. (2.7) into Eqs. (1.35-1.38). Dropping the frequency arguments of $p_{1,2}$ and $\chi_{1,2}$ in favor of a simpler notation, the boundary conditions are given by

$$p_{0,1,2}(v_T) = 0 \; , \tag{2.11}$$

$$-D\partial_v p_k(v)\Big|_{v=v_T} = \begin{cases} r_0 & \text{for} \quad k = 0 \; , \\ \chi_k & \text{for} \quad k = 1,2 \; , \end{cases} \tag{2.12}$$

$$\lim_{\Delta v \to 0}\left(p_{0,1,2}(v_R + \Delta v) - p_{0,1,2}(v_R - \Delta v)\right) = 0 \; , \tag{2.13}$$

$$\lim_{v \to -\infty} p_{0,1,2}(v) = 0 \; . \tag{2.14}$$

Note, that due to the suitable choice of the ansatz in Eq. (2.6) and Eq. (2.7), the hierarchy of equations Eqs. (2.8-2.14) are independent of α, β, and φ and now allow the computation of the functions r_0, χ_1, and χ_2, which are independent of the specific choice of the signal. Using the boundary conditions for p_0 in Eqs. (2.11-2.14) for the determination of the integration boundaries, Eq. (2.8) can be integrated twice to yield a solution for the zeroth order density. The unperturbed firing rate r_0 is then obtained by employing the normalization condition ($\int dv p_0(v) + r_0\tau_r = 1$). The result is already

well-known [73, 74] and reads

$$r_0 = \left[\tau_r + \sqrt{\pi}\int_{(\mu-v_R)/\sqrt{2D}}^{(\mu-v_T)/\sqrt{2D}} dx\; e^{x^2}\,\mathrm{erfc}(x)\right]^{-1} . \tag{2.15}$$

Now we will solve the first- and second-order equations via the Green's function method [79, 80]. The Green's function for Eq. (2.9) and Eq. (2.10) is given by the solution of

$$0 = (\hat{L}_0 + i\omega)p_h(v, v_R) + \delta(v - v_R) \tag{2.16}$$

and the boundary conditions

$$p_h(v_T, v_R) = 0 , \tag{2.17}$$

$$\lim_{\Delta v\to 0}\left(p_h(v_R + \Delta v, v_R) - p_h(v_R - \Delta v, v_R)\right) = 0 , \tag{2.18}$$

$$\lim_{v\to-\infty} p_h(v, v_R) = 0 . \tag{2.19}$$

Note, that in order to use the solution of Eq. (2.16) for the Green's function method later, we expressed p_h as a function of two arguments, v and v_R. Now we rewrite the density p_h in terms of an exponential function and an auxiliary function q_h:

$$p_h(v, v_R) = e^{-\frac{(v-\mu)^2-(v_R-\mu)^2}{4D}} q_h(v, v_R) . \tag{2.20}$$

Inserting Eq. (2.20) into Eq. (2.16) results in

$$0 = \left(D\partial_v^2 + \frac{1}{2} + i\omega - \frac{(v-\mu)^2}{4D}\right) q_h(v, v_R) + e^{\frac{(v-\mu)^2-(v_R-\mu)^2}{4D}}\delta(v - v_R) , \tag{2.21}$$

which is equivalent to

$$0 = \left(D\partial_v^2 + \frac{1}{2} + i\omega - \frac{(v-\mu)^2}{4D}\right) q_h(v, v_R) + \delta(v - v_R) . \tag{2.22}$$

For $v > v_R$ and $v < v_R$ the solution q_h of the above equation has to satisfy

$$0 = \left(D\partial_v^2 + \frac{1}{2} + i\omega - \frac{(v-\mu)^2}{4D}\right) q_h(v, v_R) . \tag{2.23}$$

The delta function at $v = v_R$ from Eq. (2.22) is then incorporated via the boundary condition

$$\lim_{\Delta v\to 0}\left(q_h'(v_R + \Delta v, v_R) - q_h'(v_R - \Delta v, v_R)\right) = 1 . \tag{2.24}$$

Further boundary conditions for q_h are obtained by inserting Eq. (2.20) into Eqs. (2.17-2.19) and read

$$q_h(v_T, v_R) = 0 \,, \tag{2.25}$$

$$\lim_{\Delta v \to 0} \left(q_h(v_R + \Delta v, v_R) - q_h(v_R - \Delta v, v_R)\right) = 0 \,, \tag{2.26}$$

$$\lim_{v \to -\infty} q_h(v, v_R) = 0 \,. \tag{2.27}$$

The solution to Eq. (2.23) and to the boundary conditions Eqs. (2.25-2.27) is given by

$$q_h(v, v_R) = \frac{1}{D d_1(v_T)} \begin{cases} \left[d_1(v) d_2(v_T) - d_1(v_T) d_2(v)\right] d_1(v_R) & \text{for} \quad v > v_R \,, \\ \left[d_1(v_R) d_2(v_T) - d_1(v_T) d_2(v_R)\right] d_1(v) & \text{for} \quad v < v_R \,. \end{cases} \tag{2.28}$$

The functions d_1 and d_2 from Eq. (2.28) are the parabolic cylinder functions [76]

$$d_1(v) = \mathcal{D}_{i\omega}\Big(\frac{\mu - v}{\sqrt{D}}\Big) \quad \text{and} \quad d_2(v) = A \cdot \mathcal{D}_{i\omega}\Big(-\frac{\mu - v}{\sqrt{D}}\Big) \tag{2.29}$$

and are linearly independent solutions of

$$0 = \left(D\partial_v^2 + \frac{1}{2} + i\omega - \frac{(v-\mu)^2}{4D}\right) d_{1,2}(v) \,. \tag{2.30}$$

The function d_1 satisfies

$$\lim_{v \to -\infty} d_1(V) = 0 \,, \tag{2.31}$$

which ensures that Eq. (2.28) satisfies Eq. (2.27). It can be verified by insertion that Eq. (2.28) also satisfies Eq. (2.25) and Eq. (2.26). In order for Eq. (2.28) to also satisfy Eq. (2.24) we have to choose the prefactor A in Eq. (2.29) appropriately. Therefore, first note that it can be shown by means of Abel's identity [81, p. 132] that the Wronskian of the solutions of Eq. (2.30) is constant. The value of this constant will depend on the prefactor A in Eq. (2.29). In order for Eq. (2.28) to satisfy Eq. (2.24) we need to choose the prefactor A such that the Wronskian equals to one, i.e.

$$d_1'(v) d_2(v) - d_1(v) d_2'(v) = 1 \,. \tag{2.32}$$

Interestingly, the exact form of A is not required for the further analysis. All we need to know is that A can be chosen such that the relation in Eq. (2.32) holds. The relation in Eq. (2.32) can also be used to simplify the derivative of q_h at the threshold

$$\begin{aligned} \partial_v q_h(v, v_R)\Big|_{v=v_T} &= \frac{1}{D d_1(v_T)} \left[d_1'(v_T) d_2(v_T) - d_1(v_T) d_2'(v_T)\right] d_1(v_R) \\ &= \frac{d_1(v_R)}{D d_1(v_T)} \,. \end{aligned} \tag{2.33}$$

Using Eq. (2.33) together with Eq. (2.25) we find for the derivative of p_h in Eq. (2.20)

$$\partial_v p_h(v, v_R)\Big|_{v=v_T} = \frac{e^{-\frac{(v_T-\mu)^2-(v_R-\mu)^2}{4D}}}{D}\frac{d_1(v_R)}{d_1(v_T)} \,. \tag{2.34}$$

Now, employing the Green's function method, we can use p_h for the solution of Eq. (2.9) and Eq. (2.10), which then read

$$p_1(v) = \chi_1 e^{i\omega\tau_r} p_h(v, v_R) - \int_{-\infty}^{v_T} dv'\, p_h(v, v')\Big[\partial_v p_0(v)\Big]\Big|_{v=v'} \tag{2.35}$$

and

$$p_2(v) = \chi_2 e^{i(\omega_1+\omega_2)\tau_r} p_h(v, v_R) - \frac{1}{2}\int_{-\infty}^{v_T} dv'\, p_h(v, v')\Big[\partial_v p_1(v, \omega_1) + \partial_v p_1(v, \omega_2)\Big]\Big|_{v=v'} \,, \tag{2.36}$$

respectively. In order to arrive at explicit expressions for the response functions χ_1 and χ_2 we take the derivative with respect to v evaluated at the point $v = v_T$ on both sides of Eq. (2.35) and Eq. (2.36) and arrive at

$$\partial_v p_1(v)\Big|_{v=v_T} = \chi_1 e^{i\omega\tau_r}\partial_v p_h(v, v_R)\Big|_{v=v_T} - \int_{-\infty}^{v_T} dv'\, \Big[\partial_v p_h(v, v')\Big]\Big|_{v=v_T}\Big[\partial_v p_0(v)\Big]\Big|_{v=v'} \,, \tag{2.37}$$

$$\begin{aligned}\partial_v p_2(v)\Big|_{v=v_T} &= \chi_2 e^{i(\omega_1+\omega_2)\tau_r}\partial_v p_h(v, v_R)\Big|_{v=v_T} \\ &\quad - \frac{1}{2}\int_{-\infty}^{v_T} dv'\, \Big[\partial_v p_h(v, v')\Big]\Big|_{v=v_T}\Big[\partial_v p_1(v, \omega_1) + \partial_v p_1(v, \omega_2)\Big]\Big|_{v=v'} \,.\end{aligned} \tag{2.38}$$

Inserting Eq. (2.34) and the expressions for $k = 1$ and $k = 2$ from Eq. (2.12) into Eq. (2.37) and Eq. (2.38), and solving for χ_1 and χ_2, respectively, we find the explicit solutions for the first- and second-order response functions

$$\chi_1(\omega) = (-1)\frac{\int_{-\infty}^{v_T} dv\, p_0'(v) e^{\frac{(v-\mu)^2-(v_T-\mu)^2}{4D}}\mathcal{D}_{i\omega}\Big(\frac{\mu-v}{\sqrt{D}}\Big)}{\mathcal{D}_{i\omega}\Big(\frac{\mu-v_T}{\sqrt{D}}\Big) - e^{i\omega\tau_r}e^{\frac{(v_R-\mu)^2-(v_T-\mu)^2}{4D}}\mathcal{D}_{i\omega}\Big(\frac{\mu-v_R}{\sqrt{D}}\Big)} \tag{2.39}$$

and

$$\chi_2(\omega_1,\omega_2) = -\frac{1}{2}\frac{\int_{-\infty}^{v_T} dv\ \left[p_1'(v,\omega_1)+p_1'(v,\omega_2)\right] e^{\frac{(v-\mu)^2-(v_T-\mu)^2}{4D}} \mathcal{D}_{i(\omega_1+\omega_2)}\left(\frac{\mu-v}{\sqrt{D}}\right)}{\mathcal{D}_{i(\omega_1+\omega_2)}\left(\frac{\mu-v_T}{\sqrt{D}}\right) - e^{i(\omega_1+\omega_2)\tau_r} e^{-\frac{(v_R-\mu)^2-(v_T-\mu)^2}{4D}} \mathcal{D}_{i(\omega_1+\omega_2)}\left(\frac{\mu-v_R}{\sqrt{D}}\right)} \ . \tag{2.40}$$

The expressions in Eq. (2.39) and Eq. (2.40) can be simplified further by defining

$$f(v,\omega) = e^{\frac{(v-\mu)^2-(v_T-\mu)^2}{4D}} \mathcal{D}_{i\omega}\left(\frac{\mu-v}{\sqrt{D}}\right) \tag{2.41}$$

and noting that the function f in Eq. (2.41) satisfies the following relation

$$0 = Df''(v,\omega) - (v-\mu)f'(v,\omega) + i\omega f(v,\omega)\ . \tag{2.42}$$

The validity of Eq. (2.42) can be verified by insertion of Eq. (2.41) into Eq. (2.42) and by using that the parabolic cylinder function is a solution of Eq. (2.30). By partial integration, employing the differential equation Eq. (2.8) with the respective boundary conditions, using that the functions p_0 and f decay sufficiently fast for $v \to -\infty$, and employing the relation Eq. (2.42), we find

$$\int_{-\infty}^{v_T} dv\ p_0'(v) f(v,\omega) = \frac{r_0}{1-i\omega}\left(f'(v_R,\omega) - f'(v_T,\omega)\right)\ . \tag{2.43}$$

Equivalently, by employing the differential equation Eq. (2.9) with the respective boundary conditions, and using that the function p_1 decays sufficiently fast for $v \to -\infty$, we find

$$\begin{aligned}\int_{-\infty}^{v_T} dv\ p_1'(v,\omega_{1,2}) f(v,\omega_1+\omega_2) = {}& \frac{\chi_1(\omega_{1,2})\left(f'(v_R,\omega_1+\omega_2)e^{i\omega_1\tau_r} - f'(v_T,\omega_1+\omega_2)\right)}{1+i(\omega_{1,2}-\omega_1-\omega_2)} \\ & + \frac{1}{1+i(\omega_{1,2}-\omega_1-\omega_2)} \int_{-\infty}^{v_T} dv\ p_0'(v) f'(v,\omega_1+\omega_2)\ ,\end{aligned} \tag{2.44}$$

and

$$\int_{-\infty}^{v_T} dv\ p_0'(v) f'(v,\omega_1+\omega_2) = \frac{r_0}{2-i(\omega_1+\omega_2)}\left(f''(v_R,\omega) - f''(v_T,\omega)\right)\ . \tag{2.45}$$

Utilizing that for the parabolic cylinder function [76]

$$\partial_v \mathcal{D}_{i\omega-l}\left(\frac{\mu-v}{\sqrt{D}}\right) + \frac{v-\mu}{2D}\mathcal{D}_{i\omega-l}\left(\frac{\mu-v}{\sqrt{D}}\right) = \frac{-i\omega-l}{\sqrt{D}}\mathcal{D}_{i\omega-l-1}\left(\frac{\mu-v}{\sqrt{D}}\right)\ , \tag{2.46}$$

we can express the derivative of the function $f(v,\omega)$ in Eq. (2.41) as

$$\partial_v f(v,\omega) = e^{\frac{(v-\mu)^2-(v_T-\mu)^2}{4D}} \mathcal{D}_{i\omega-1}\left(\frac{\mu-v}{\sqrt{D}}\right)\frac{(-i\omega)}{\sqrt{D}} \tag{2.47}$$

and the second derivative as

$$\partial_v^2 f(v,\omega) = e^{\frac{(v-\mu)^2-(v_T-\mu)^2}{4D}} \mathcal{D}_{i\omega-2}\left(\frac{\mu-v}{\sqrt{D}}\right)\frac{(-i\omega)(-i\omega-1)}{D}\,. \tag{2.48}$$

Using Eqs. (2.43-2.45), Eq. (2.47) and Eq. (2.48) it is now possible to simplify Eq. (2.39) and Eq. (2.40). For the first-order response function we obtain

$$\chi_1(\omega) = r_0 \frac{i\omega/\sqrt{D}}{i\omega-1}\frac{\mathcal{D}_{i\omega-1}(\frac{\mu-v_T}{\sqrt{D}}) - e^{\Delta}\mathcal{D}_{i\omega-1}(\frac{\mu-v_R}{\sqrt{D}})}{\mathcal{D}_{i\omega}(\frac{\mu-v_T}{\sqrt{D}}) - e^{\Delta}e^{i\omega\tau_r}\mathcal{D}_{i\omega}(\frac{\mu-v_R}{\sqrt{D}})} \tag{2.49}$$

with $\Delta = [v_R^2 - v_T^2 + 2\mu(v_T - v_R)]/(4D)$. The same result as in Eq. (2.49), although by means of a different kind of calculation, has been first obtained by [41] and [42], where the authors in the latter study employed a different notation than in Eq. (2.49). The equivalence of the results for χ_1 from [41] and [42] has been demonstrated in [82]. An interesting alternative derivation of χ_1 exploits an analogy of the Fokker-Planck equation, Eq. (1.34), with the Hamiltonian of a quantum harmonic oscillator [83]. Here, we derived Eq. (2.49) in order to demonstrate that our derivation method indeed reproduces known and verified results.

Now we arrive at the main result of this chapter - the second-order response function χ_2. After simplification of Eq. (2.40) by means of Eqs. (2.44-2.48) we find

$$\begin{aligned}\chi_2(\omega_1,\omega_2) =& r_0\frac{(1-i\omega_1-i\omega_2)(i\omega_1+i\omega_2)}{2D(i\omega_1-1)(i\omega_2-1)}\frac{\mathcal{D}_{i\omega_1+i\omega_2-2}(\frac{\mu-v_T}{\sqrt{D}}) - e^{\Delta}\mathcal{D}_{i\omega_1+i\omega_2-2}(\frac{\mu-v_R}{\sqrt{D}})}{\mathcal{D}_{i\omega_1+i\omega_2}(\frac{\mu-v_T}{\sqrt{D}}) - e^{\Delta}e^{i(\omega_1+\omega_2)\tau_r}\mathcal{D}_{i\omega_1+i\omega_2}(\frac{\mu-v_R}{\sqrt{D}})} \\ &+ \frac{i\omega_1+i\omega_2}{2\sqrt{D}}\frac{\left(\frac{\chi_1(\omega_1)}{i\omega_2-1}+\frac{\chi_1(\omega_2)}{i\omega_1-1}\right)\mathcal{D}_{i\omega_1+i\omega_2-1}(\frac{\mu-v_T}{\sqrt{D}})}{\mathcal{D}_{i\omega_1+i\omega_2}(\frac{\mu-v_T}{\sqrt{D}}) - e^{\Delta}e^{i(\omega_1+\omega_2)\tau_r}\mathcal{D}_{i\omega_1+i\omega_2}(\frac{\mu-v_R}{\sqrt{D}})} \\ &- \frac{i\omega_1+i\omega_2}{2\sqrt{D}}\frac{\left(\frac{\chi_1(\omega_1)e^{i\omega_1\tau_r}}{i\omega_2-1}+\frac{\chi_1(\omega_2)e^{i\omega_2\tau_r}}{i\omega_1-1}\right)e^{\Delta}\mathcal{D}_{i\omega_1+i\omega_2-1}(\frac{\mu-v_R}{\sqrt{D}})}{\mathcal{D}_{i\omega_1+i\omega_2}(\frac{\mu-v_T}{\sqrt{D}}) - e^{\Delta}e^{i(\omega_1+\omega_2)\tau_r}\mathcal{D}_{i\omega_1+i\omega_2}(\frac{\mu-v_R}{\sqrt{D}})}\,.\end{aligned} \tag{2.50}$$

Together with Eq. (2.15) and Eq. (2.49) the second-order response function can now be utilized for the computation of the weakly nonlinear response, Eq. (2.1), for arbitrary time-dependent signals.

The amplitude and phase of χ_2 for two firing regimes of the LIF model are shown in Fig. 2.1. The first regime is the subthreshold firing regime (Fig. 2.1a,b) in which the mean current μ is smaller than the threshold. In this regime noise is required to reach the threshold and spike trains are typically irregular with a high coefficient of variation (CV) [84]. For our choice of parameters we find $r_0 \approx 0.14$ and $CV \approx 0.4$. In the subthreshold

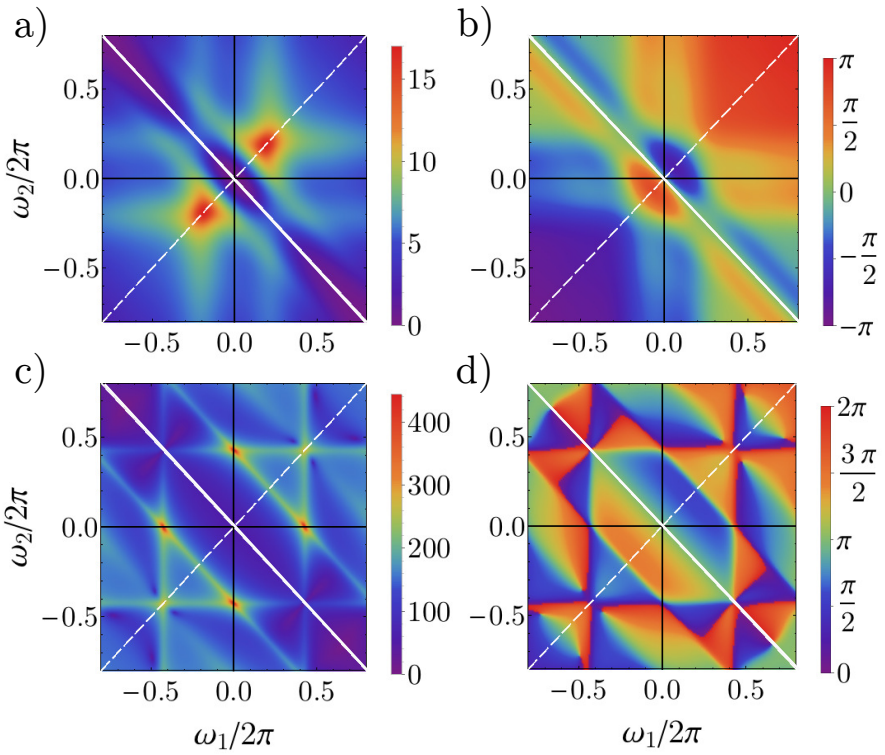

Figure 2.1: Second-order response functions for the LIF model. Amplitude (a,c) and phase (b,d) of the second-order response kernel, Eq. (2.50), for the LIF model in the subthreshold regime (a,b) and the suprathreshold regime (c,d). For the subthreshold regime we used $\mu = 0.9$, $D = 0.005$. For the suprathreshold regime we employed $\mu = 1.1$, $D = 0.001$.

regime, the amplitude of χ_2 exhibits two broad peaks (Fig. 2.1a). The second regime is the suprathreshold regime with a mean current μ which is higher than the threshold. In this regime, the model exhibits sustained firing in the absence of noise and the spike trains are more regular than in the subthreshold regime resulting in a small CV [84]. In our case we find $r_0 \approx 0.42$ and $CV \approx 0.02$. The amplitude of χ_2 for the suprathreshold regime displays triangular structure (Fig. 2.1c) in marked contrast to the subthreshold case. From Eq. (2.6) it can be seen that upon exchange of the frequencies ω_1 and ω_2 the dynamics of the response does not change. Consequently, the function χ_2 obeys the symmetry

$$\chi_2(\omega_1, \omega_2) = \chi_2(\omega_2, \omega_1) \, . \tag{2.51}$$

Furthermore, from the definition of the double Fourier transform of the kernel K_2 in

Eq. (2.3) we find

$$\chi_2(\omega_1, \omega_2) = \chi_2^*(-\omega_1, -\omega_2) \,, \tag{2.52}$$

where the asterisk denotes the complex conjugated. Combining Eq. (2.51) and Eq. (2.52) it follows for the imaginary part of $\chi_2(\omega, -\omega)$ that

$$\mathrm{Im}\left(\chi_2(\omega, -\omega)\right) = 0 \,. \tag{2.53}$$

The symmetries from Eq. (2.51) and Eq. (2.52) become apparent in Fig. 2.1 and imply that the second, third and fourth quadrants of the graphs can be inferred from the first quadrant. The property in Eq. (2.53) becomes manifest in the fact, that the phase of χ_2 along the solid white diagonal in Fig. 2.1b,d is either π or zero.

Interpretation of the second-order response function

In order to interpret the structures in Fig. 2.1 and the meaning of the second-order response function for different frequency arguments, we again consider a signal consisting of a sum of two cosines,

$$s(t) = \alpha \cos(\omega_1 t) + \beta \cos(\omega_2 t + \varphi) \,, \tag{2.54}$$

where α and β are constants and φ is the relative phase between the two cosines. Above, it was already shown that the response to Eq. (2.54) is given by Eq. (2.6). Rewriting the complex exponentials in Eq. (2.6) in terms of cosine functions, we arrive at

$$\begin{aligned} r(t) \approx\ & \underline{r_0 + \frac{\varepsilon^2\alpha^2}{2}\chi_2(\omega_1, -\omega_1) + \frac{\varepsilon^2\beta^2}{2}\chi_2(\omega_2, -\omega_2)} \\ & + \varepsilon\Big[\ \alpha\,|\chi_1(\omega_1)|\cos[\omega_1 t - \phi_1(\omega_1)] + \beta\,|\chi_1(\omega_2)|\cos[\omega_2 t + \varphi - \phi_1(\omega_2)]\ \Big]_{LR} \\ & + \frac{\varepsilon^2}{2}\Big(\ \alpha^2\,|\chi_2(\omega_1, \omega_1)|\cos[2\omega_1 t - \phi_2(\omega_1, \omega_1)] \\ & + \beta^2\,|\chi_2(\omega_2, \omega_2)|\cos[2\omega_2 t + 2\varphi - \phi_2(\omega_2, \omega_2)]\ \Big)_{HH} \\ & + \varepsilon^2\alpha\beta\Big\{\ |\chi_2(\omega_1, \omega_2)|\cos[(\omega_1 + \omega_2)t + \varphi - \phi_2(\omega_1, \omega_2)] \\ & + |\chi_2(\omega_1, -\omega_2)|\cos[(\omega_1 - \omega_2)t - \varphi - \phi_2(\omega_1, -\omega_2)]\ \Big\}_{MR} \,, \end{aligned} \tag{2.55}$$

where ϕ_1 and ϕ_2 are the complex arguments of χ_1 and χ_2, respectively. Whenever we employ the signal in Eq. (2.54) we will refer to Eq. (2.55) as the weakly nonlinear theory. Note, that Eq. (2.54) is not specific to the LIF model but is valid for the weakly nonlinear response of any signal-processing system that is driven by a sum of two cosine signals. The specific behavior of a given model is only incorporated after specifying r_0, χ_1 and χ_2. Consequently, our following interpretations of the first- and second-order response functions are not restricted to the LIF model but are of a broader scope.

The terms in Eq. (2.55) inside of the square brackets which are denoted by $[...]_{LR}$ correspond to the linear response and are proportional to the linear response function χ_1. For cosine signals the linear response comprises oscillations which oscillate with the same frequency as the cosines in the input signal (ground modes). For general time-dependent signals which can be decomposed into Fourier modes, the linear response corresponds to contributions of the Fourier modes to oscillations in the output which oscillate with the same frequency as the Fourier modes of the input signal and whose amplitude and phase is determined by the function χ_1. In the following, whenever we employ a signal like in Eq. (2.54) and refer to the linear response theory, we mean Eq. (2.55) with only the zero-order and first-order terms in ε.

The terms in Eq. (2.55) which are inside of the round brackets $(...)_{HH}$ comprise the *higher harmonics* in the output. The higher harmonics are oscillations which oscillate at doubled frequencies (2ω) with respect to the initial input and whose amplitude and phase is determined by $\chi_2(\omega, \omega)$. For time-dependent signals, $\chi_2(\omega, \omega)$ determines the amplitude and phase of oscillations with frequencies 2ω which are excited by the Fourier mode of the signal at frequency ω. The values of the amplitude and phase of $\chi_2(\omega, \omega)$ are plotted along the white dashed diagonals in Fig. 2.1. For the subthreshold regime in Fig. 2.1a,b, the strongest higher harmonic is excited by an input oscillation of $\omega/2\pi \approx 0.2$, which is a little bit less than twice the firing rate ($r_0 \approx 0.14$ in the subthreshold regime). For the suprathreshold regime in Fig. 2.1c,d, the strongest higher harmonic is excited by an oscillation with $\omega/(2\pi) = r_0/2 \approx 0.21$ and $\omega/(2\pi) = r_0 \approx 0.42$.

The terms in Eq. (2.55) inside the curly brackets $\{...\}_{MR}$ correspond to the *mixed response* and their amplitude and phase are determined by $\chi_2(\omega_1, \pm\omega_2)$. The mixed response is comprised of oscillations with frequencies $\omega_1 + \omega_2$ and $|\omega_1 - \omega_2|$, i.e. of oscillations with frequencies that correspond to sums or differences of the signal frequencies. For time-dependent signals, the function $\chi_2(\omega_1, \pm\omega_2)$ determines the amplitude and phase of oscillations which arise due to interactions between two different Fourier modes of the input signal. The amplitude and phase of $\chi_2(\omega_1, \omega_2)$ which are plotted off the white dashed diagonals in Fig. 2.1 indicate how strong the mixed response in the LIF model is excited by a combination of input oscillations with frequencies ω_1 and ω_2. For example, in the suprathreshold regime, which is shown in Fig. 2.1c,d, the mixed response is strongest for input oscillations with $(\omega_1 + \omega_2)/2\pi = r_0 \approx 0.42$, while in the subthreshold regime in Fig. 2.1a,b the mixed response is most pronounced for $\omega_1/2\pi = \omega_2/2\pi \approx 0.21$.

The underlined terms in Eq. (2.55) correspond to the *time-averaged firing rate* because only the underlined terms would contribute to the time-average of Eq. (2.55) over a sufficiently long time window, while all time-dependent terms would be averaged out. The time-averaged firing rate consists of r_0 and second-order contributions which are proportional to $\chi_2(\omega, -\omega)$. The linear response does not contribute to any time-independent terms in Eq. (2.55). For general time-dependent signals, the function $\chi_2(\omega, -\omega)$ determines how strongly a Fourier mode of the signal at frequency ω contributes to the time-average of the firing rate. The amplitude of $\chi_2(\omega, -\omega)$, which is plotted along the white solid diagonals in Fig. 2.1a,c, indicates how strong the time-averaged firing rate is

altered by an input oscillation with frequency ω, while the phase of $\chi_2(\omega,\omega)$, which is plotted along the white solid diagonals in Fig. 2.1b,d, indicates whether the alteration is positive or negative. From the phase of χ_2 along the white solid diagonal for the subthreshold regime in Fig. 2.1b we can see that the phase is always zero and the alterations of the time-averaged firing rate by input oscillations of frequency ω are always positive. For the suprathreshold regime, however, we find that the phase of χ_2 along the white solid diagonal in Fig. 2.1d leads to a negative sign of the alteration of the time-averaged firing rate for small ω and that the alteration changes sign for larger ω.

In order to obtain an interpretation for χ_2 along the frequency axes (black solid lines in Fig. 2.1) and also along the white dashed diagonal, we now consider a signal consisting of a sum of a fast cosine and a slow cosine, i.e. we consider

$$s(t) = \alpha\cos(\omega_{fast}t) + \beta(\cos(\omega_{slow}t+\varphi) \tag{2.56}$$

where we take $\omega_{slow}\to 0$. A system which is driven by the signal in Eq. (2.56) can be also viewed as a system which is effectively driven by a single cosine (the first term in Eq. (2.56)) but for which the mean current μ in the Langevin equation, Eq. (1.33), is replaced by the time-dependent current

$$\mu_s(t) = \mu + \varepsilon\beta(\cos(\omega_{slow}t+\varphi)\ . \tag{2.57}$$

The second-order response of a system with a time-dependent current $\mu_s(t)$ and a signal which consists of one cosine reads

$$\begin{aligned} r(t) \approx\ & r_0(\mu_s(t)) + \frac{\varepsilon\alpha}{2}\Big(\chi_1(\mu_s(t),\omega_{fast})e^{-i\omega_{fast}t} + c.c\Big) \\ & + \frac{\varepsilon^2\alpha^2}{4}\Big(\chi_2(\mu_s(t),\omega_{fast},\omega_{fast})e^{-i2\omega_{fast}t} + c.c.\Big) \\ & + \frac{\varepsilon^2\alpha^2}{2}\chi_2(\mu_s(t),\omega_{fast},-\omega_{fast})\ , \end{aligned} \tag{2.58}$$

where the function r_0, χ_1 and χ_2 are now evaluated for the time-dependent current $\mu_s(t)$. Now, employing Eq. (2.57), expanding Eq. (2.58) with respect to ε, and omitting terms of third order in ε and higher we find

$$\begin{aligned} r(t) \approx\ & r_0(\mu) + \varepsilon\alpha|\chi_1(\omega_{fast})|\cos[\omega_{fast}t - \phi_1(\omega_{fast})] + \varepsilon\beta\frac{\partial}{\partial\mu}r_0(\mu)\cos[\omega_{slow}t+\varphi] \\ & + \frac{\varepsilon^2\beta^2}{4}\frac{\partial^2}{\partial\mu^2}r_0(\mu)\Big(1+\cos[2\omega_{slow}t+2\varphi]\Big) \\ & + \frac{\varepsilon^2\alpha^2}{2}\Big(|\chi_2(\omega_{fast},-\omega_{fast})| + |\chi_2(\omega_{fast},\omega_{fast})|\cos[2\omega_{fast}t+2\varphi-\phi_2(\omega_{fast},\omega_{fast})]\Big) \\ & + \frac{\varepsilon^2\alpha\beta}{2}\frac{\partial}{\partial\mu}\chi_1(\mu,\omega_{fast})\Big(e^{-i(\omega_{fast}+\omega_{slow})t}e^{-i\varphi} + e^{-i(\omega_{fast}-\omega_{slow})t}e^{i\varphi} + c.c.\Big)\ . \end{aligned} \tag{2.59}$$

From a comparison of Eq. (2.59) and Eq. (2.55) we find

$$\lim_{\omega\to 0}\chi_2(\omega,-\omega)=\frac{1}{2}\frac{\partial^2}{\partial\mu^2}r_0(\mu) \tag{2.60}$$

and

$$\lim_{\omega_2\to 0}\chi_2(\omega_1,\omega_2)=\lim_{\omega_2\to 0}\chi_2(\omega_1,-\omega_2)=\frac{1}{2}\frac{\partial}{\partial\mu}\chi_1(\mu,\omega_1)\ . \tag{2.61}$$

Consequently, the function values of χ_2 along the solid black lines in Fig. 2.1 can be related to the amplitude and phase of the first derivative of the linear response function and the low-frequency limit of $\chi_2(\omega,-\omega)$ at the intersections of two solid black lines is proportional to the second derivative of the static transfer function.

2.2 Weakly nonlinear response of the LIF model

Excitation of a higher harmonic by a cosine signal

One may ask at this point whether the weakly nonlinear theory can describe qualitatively different effects than the linear theory or whether the weakly nonlinear response is merely a correction to the linear response. In order to test whether the second-order response exhibits significant differences from the linear response we first choose a signal which consists of a single cosine with frequency ω_1 ($\alpha=1$ and $\beta=0$ in Eq. (2.54)) and consider the Fourier transform of Eq. (2.55)

$$\tilde{r}_T(\omega)=\int_0^T dt'\ e^{i\omega t'}r(t')\ . \tag{2.62}$$

For sufficiently long simulation times T, Eq. (2.62) reads

$$\tilde{r}_T(\omega)=\begin{cases}T\frac{1}{2}A_{GM} & \text{for}\quad \omega=\omega_1\ ,\\ T\frac{1}{2}A_{HH} & \text{for}\quad \omega=2\omega_1\ ,\\ Tr_0+T\frac{\varepsilon^2}{2}\chi_2(\omega_1,-\omega_1) & \text{for}\quad \omega=0\ ,\\ 0 & \text{else}\ ,\end{cases} \tag{2.63}$$

where $A_{GM}=\varepsilon\chi_1(\omega_1)$ is the amplitude of the oscillation in the output with frequency ω_1 (the ground mode) and where $A_{HH}=\varepsilon^2\chi_2(\omega_1,\omega_1)/2$ is the amplitude of the output oscillation with frequency $2\omega_1$ (the higher harmonic). Numerically, the amplitude of the ground mode and the higher harmonic can be obtained by simulating the Langevin equation for the LIF model, by Fourier transforming the numerically measured time-dependent firing rate and by evaluating the Fourier transformed rate at $\omega=\omega_1$ and at $\omega=2\omega_1$. The results of this numerical procedure are shown in figure Fig. 2.2 together with the theoretical prediction of A_{GM} and A_{HH} from Eq. (2.63).

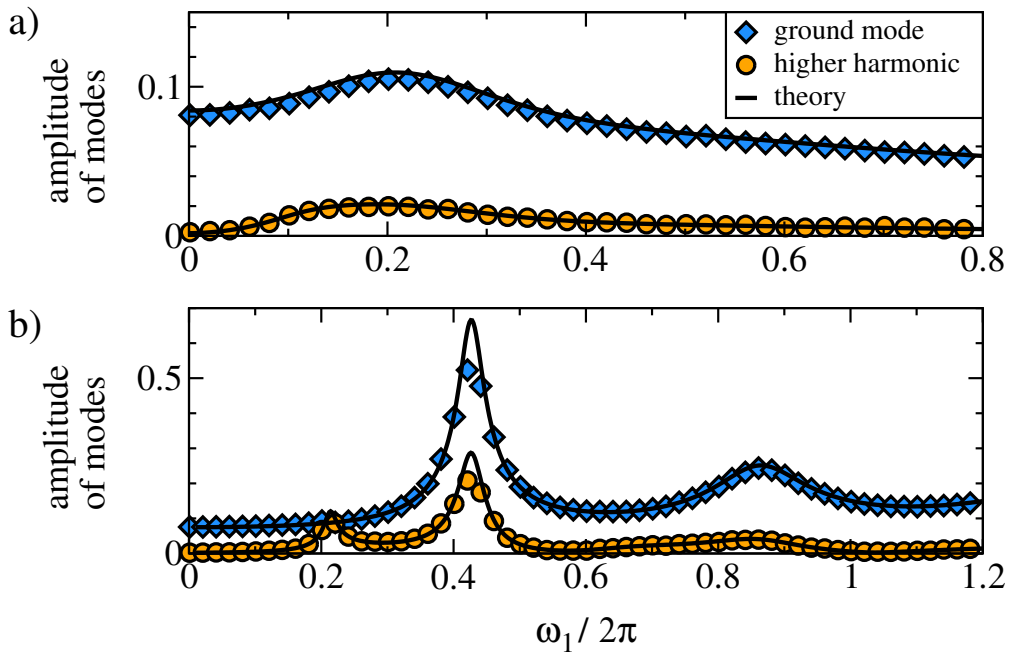

Figure 2.2: Amplitude of ground mode and amplitude of higher harmonic vs frequency of input oscillation. (a) subthreshold firing regime; (b) suprathreshold firing regime. Symbols correspond to numerical measurements of the amplitudes. The theoretical predictions of the amplitude of the ground mode and the higher harmonic are given by A_{GM} and A_{HH} from Eq. (2.63), respectively. The signal amplitude is $\varepsilon = 0.05$ and the remaining parameters are as in Fig. 2.1.

In the subthreshold regime in Fig. 2.2a the ground mode is about an order of magnitude higher than the amplitude of the higher harmonic regardless of the frequency of the driving frequency. We conclude, that in this regime the higher harmonic does not contribute much to the modulation of the firing rate and that the weakly nonlinear response merely constitutes a correction to the linear response.

In the suprathreshold regime in Fig. 2.2b, however, the relative amplitude of the higher harmonic with respect to the ground mode is strongly dependent on the signal frequency. For $\omega_1/(2\pi) \approx r_0/2 = 0.21$ the amplitude of the higher harmonic even exceeds the amplitude of the ground mode. Of course, the theoretical prediction for the amplitude of the higher harmonic in Eq. (2.63) exhibits a quadratic dependence on ε, while the ground mode is linear in ε. It may seem that increasing ε would eventually lead to a dominance of the higher harmonic in Eq. (2.63). However, increasing ε also leads to increasing contributions from third- and higher-order terms in ε, which we omitted in Eq. (2.63) and which would lead to deviations of the weakly nonlinear theory from the true LIF firing rate. A first indication of such higher-order contributions is visible for $\omega_1/(2\pi) \approx 0.42$ in Fig. 2.2b, where the weakly nonlinear theory overestimates the numerically measured amplitudes of the ground mode and of the higher harmonic. For

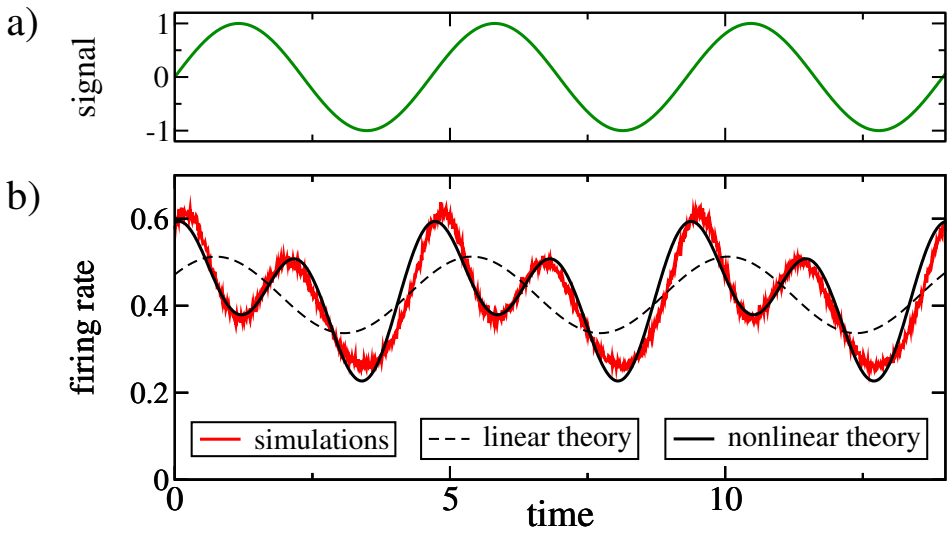

Figure 2.3: Excitation of a harmonic oscillation in the firing rate of the LIF model for the suprathreshold firing regime. (a) Cosine signal with $\alpha = 1$, $\beta = 0$, and $\omega_1/2\pi = 0.21$. (b) Time-dependent firing rate (red) from numerical simulations of the LIF model. The nonlinear theory (solid black line) is given by Eq. (2.55). The linear theory (dashed black line) is computed by Eq. (2.55), where terms proportional to ε^2 were omitted. $\varepsilon = 0.05$ and the remaining parameters are the same as in Fig. 2.1c,d.

other signal frequencies, higher-order contributions can be safely neglected for $\varepsilon = 0.05$, as indicated by the close match between the numerical measurements of the amplitudes with the second-order prediction from Eq. (2.63).

For a signal with $\alpha = 1$, $\beta = 0$, and $\omega_1/2\pi = 0.21$ we now plot in Fig. 2.3 the firing rate of the LIF model in the suprathreshold firing regime. The higher harmonic, with an oscillation frequency of $2\omega_1$, can be clearly seen in the time-dependent firing rate. The linear response theory in this case does not only provide a quantitatively poor prediction of the output but also fails to predict the position of the peaks in the response.

At this point we would like to compare the weakly nonlinear response theory with the linear-nonlinear theory (LN theory) put forward by Ostojic and Brunel [85]. The LN theory combines the linear response kernel K_1 with the unperturbed firing rate r_0, which serves as a static nonlinearity such that the time-dependent firing rate, $r(t)$, can be represented as

$$r_{\mathrm{LN}}(t) = r_0\Big(f_{\mathrm{LN}}(t)\Big) \tag{2.64}$$

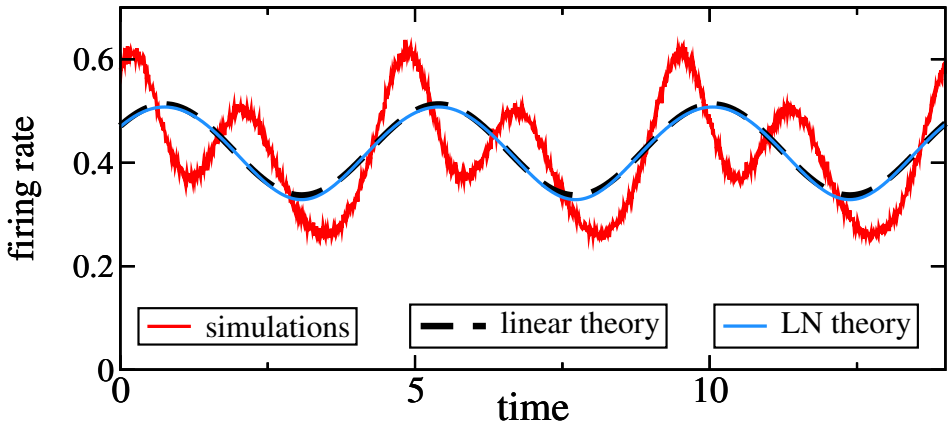

Figure 2.4: The LN theory (blue line) fails to describe the LIF firing rate (red curve) in the suprathreshold regime and does not differ significantly from the linear response theory (dashed black line). Model and signal parameters are the same as in Fig. 2.3.

where for a general time-dependent signal the function $f_{\text{LN}}(t)$ is given by

$$f_{\text{LN}}(t) = \mu + \frac{\int_{-\infty}^{\infty} dt'\, K_1(t')s(t-t')}{\frac{\partial}{\partial\mu} r_0(\mu)} \,. \tag{2.65}$$

For a single cosine with amplitude ε and signal frequency ω_1, Eq. (2.65) reduces to

$$f_{\text{LN}}(t) = \mu + \varepsilon \frac{|\chi_1(\omega_1)|}{\chi_1(0)} \cos\left(\omega_1 t - \phi_1(\omega_1)\right) \,. \tag{2.66}$$

The theory, Eq. (2.64) is set up such that it becomes exact in the limit of weak signals (in which case the linear response is valid) and in the limit of adiabatically slow signals but arbitrary signal strengths (in which case the response is governed by the static transfer function). Interestingly, the theory also works remarkably well for strong and fast signals as long the intrinsic noise of the neuron is sufficiently strong [85].

In Fig. 2.4 we plot the LN theory from Eq. (2.64) where we insert $\mu(t)$ from Eq. (2.66) together with the LIF firing rate for the same signal and the same firing regime as in Fig. 2.3. For the considered model and signal parameters, the LN theory by Ostojic and Brunel yields practically the same result as the linear response theory. This is because the static nonlinearity is essentially linear over the relevant range of signal values as can be seen from Fig. 2.5. Furthermore, the LN theory cannot account for differences in the frequency-dependence of the amplitudes of the ground mode and the higher harmonic because the frequency-dependence in the LN theory stems solely from the linear response function. This is why the LN theory fails to predict the exact time-course of the rate-modulation in Fig. 2.4. The weakly nonlinear response theory, however,

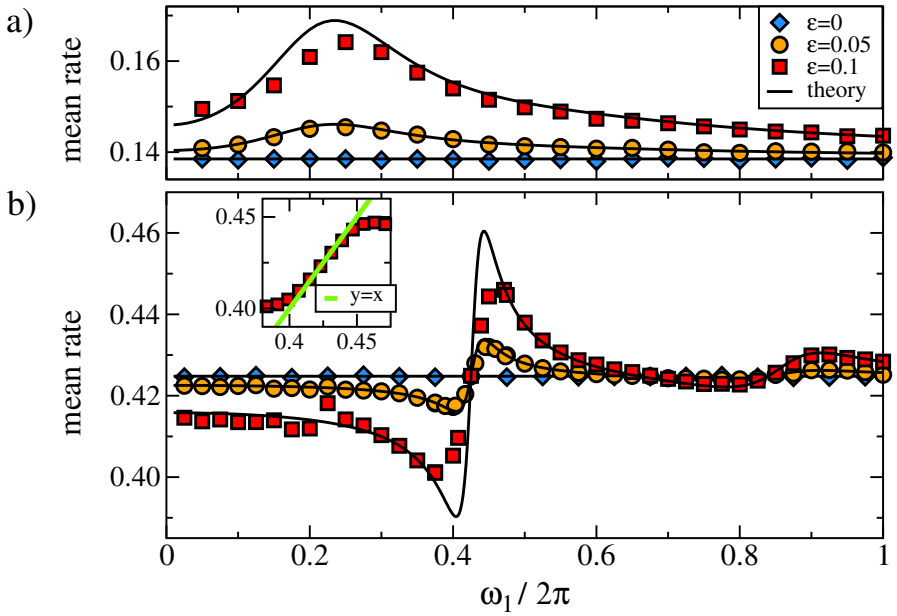

Figure 2.5: Time-averaged firing rate vs signal frequency for the LIF model for $\varepsilon = 0$ (blue diamonds), $\varepsilon = 0.05$ (orange circles), and $\varepsilon = 0.1$ (red squares). a) Subthreshold regime; b) Suprathreshold regime; Parameters were the same as in Fig. 2.1.

which is rigorously derived from the true dynamics of the model correctly accounts for the frequency-dependence of the second-order terms in Eq. (2.55), which contribute to the time-dependent firing rate.

Nonlinear effect of a cosine signal on the time-averaged firing rate

Averaging the response in Eq. (2.55) over a long time window yields the time-averaged firing rate

$$\begin{aligned}\langle r(t)\rangle_T &= \lim_{T\to\infty}\frac{1}{T}\int_0^T dt\, r(t)\\ &\approx r_0+\frac{\varepsilon^2\alpha^2}{2}\chi_2(\omega_1,-\omega_1)+\frac{\varepsilon^2\beta^2}{2}\chi_2(\omega_2,-\omega_2)\ .\end{aligned} \tag{2.67}$$

The linear response does not show up in the above equation and the first terms which describe the influence of the signal are of second-order in ε. The time-averaged firing rate in response to a single cosine ($\alpha = 1$, $\beta = 0$ in Eq. (2.54)) is plotted in Fig. 2.6 for the two firing regimes of the LIF model.

In the subthreshold regime the signal always leads to an increase of the time-averaged firing rate. For high frequencies, the increase of the rate is only small, as expected from

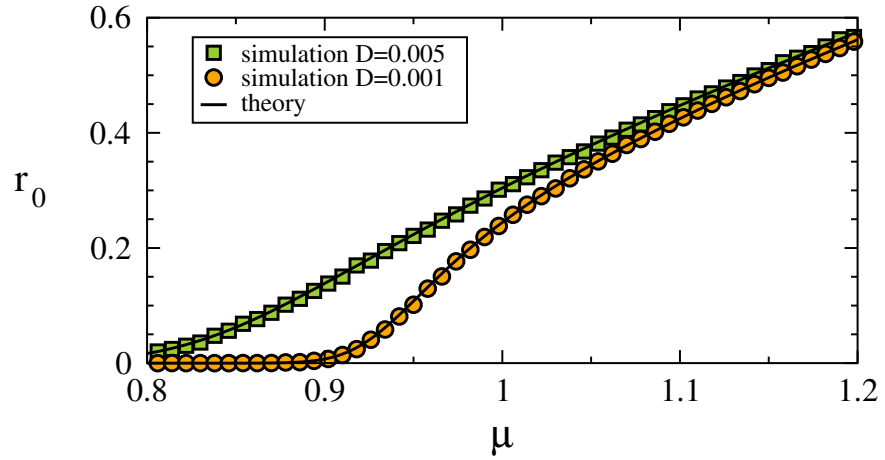

Figure 2.6: Unperturbed firing rate vs static input current for the LIF model for $D = 0.005$ (green squares) and $D = 0.001$ (orange circles). For $\mu = 0.9$ the curvature of the function $r_0(\mu)$ is positive and for $\mu = 1.1$ the curvature is negative. The noise affects the strength of the curvature at $\mu = 0.9$ and $\mu = 1.1$ but not its sign. Remaining parameters: $\varepsilon = 0$ and $\tau_r = 0$.

the low-pass nature of the LIF model. For low signal frequencies, the increase of the rate is proportional to the second derivative of the static transfer function with respect to the mean current (cf. Eq. (2.60)) such that for $\alpha = 1$ and $\beta = 0$ in Eq. (2.54) the time-averaged rate reads

$$\lim_{\omega_1 \to 0} \langle r(t) \rangle_T = r_0 + \frac{\varepsilon^2}{4} \frac{\partial^2}{\partial \mu^2} r_0(\mu) \, . \tag{2.68}$$

The function $r_0(\mu)$ is shown in Fig. 2.5 for two different noise values.

In the subthreshold regime, for $\mu = 0.9$ and $D = 0.005$, the second derivative of $r_0(\mu)$ with respect to μ is positive, which leads to an increase of the time-averaged firing rate for slow signals. For faster signals the increase of the time-averaged rate in Fig. 2.6a attains a maximum at a frequency which is larger than the unperturbed firing rate and depends on the specific choice of μ and D.

In the suprathreshold regime, a signal can lead to an increase but also to a decrease of the time-averaged firing rate depending on the signal frequency. Fast signals do not change the time-average firing rate much, just like in the subthreshold regime. Slow signals however, lead to a decrease of the time-averaged firing rate in the suprathreshold regime, because the second derivative of the static transfer function is negative for $\mu = 1.1$ (see Fig. 2.5 for $\mu = 1.1$ and $D = 0.001$). For signal frequencies which are close to the unperturbed firing rate the time-averaged firing rate exhibits a linear relationship with the signal frequency (see inset of Fig. 2.6b). A signal with a frequency which is slightly larger than r_0 will speed up the firing, while a signal with a frequency which is slightly smaller than r_0 will slow down the firing. This behavior of the LIF model is

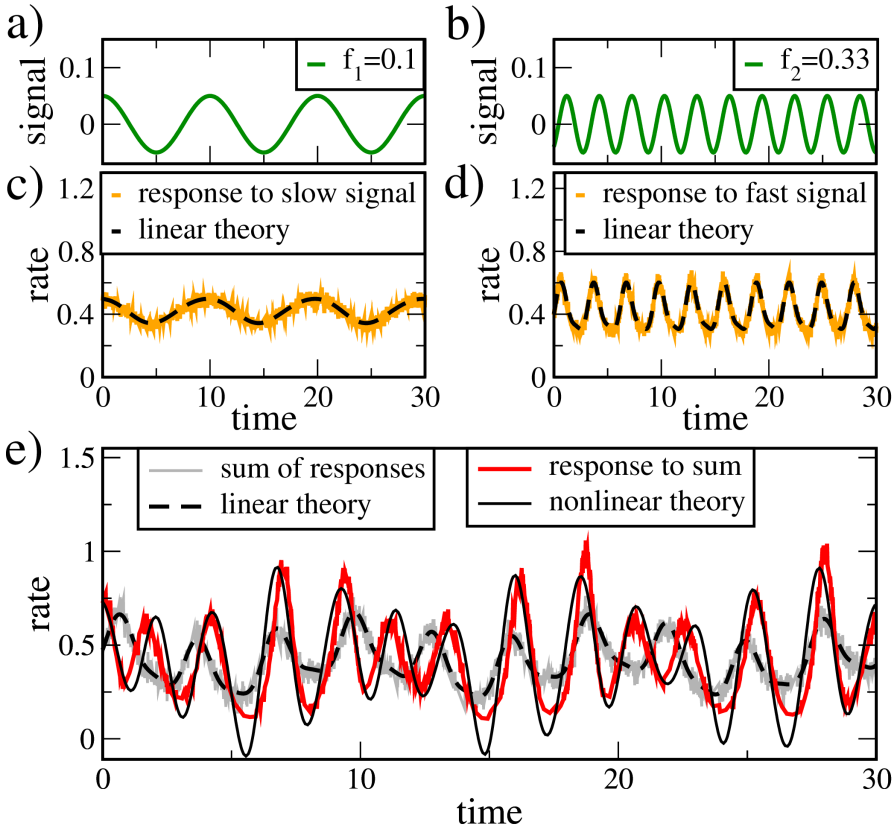

Figure 2.7: Mixed response of the LIF model in the suprathreshold regime to the sum of two cosines. a) Slow signal with $\alpha = 1$, $\beta = 0$, $f_1 = \omega_1/(2\pi) = 0.1$ and $\varepsilon = 0.05$; b) Fast signal with $\alpha = 0$, $\beta = 1$, $f_2 = \omega_2/(2\pi) = 0.33$ and $\varepsilon = 0.05$; c) The response to the slow signal from a) (orange curve) is compared to the linear response theory (dashed black line). d) The response to the fast signal from b) (orange curve) is compared to the linear response theory (dashed black line). e) The response to the sum of the slow and fast signals (red curve) is correctly predicted by the weakly nonlinear theory (solid black line) but not by the linear response (dashed black line). Due to the nonlinear behavior of the response, the response to the sum of signals (red curve) is markedly different from the sum of the responses from c) and d) (corrected for the mean firing rate). Model parameters are the same as in Fig. 2.1c,d.

reminiscent of frequency locking which has been studied in the deterministic LIF [86] and in the stochastic version of the LIF for strong input signals [87]. It is interesting, that some features of this highly nonlinear locking behavior are captured by our weakly nonlinear theory, which only takes into account second-order contributions of the signal.

Mixed response of the LIF model to a sum of cosines

So far we have studied the nonlinear response of the LIF model to one single cosine. Here, we consider the nonlinear response to a signal consisting of two cosines ($\alpha = 1$, $\beta = 1$ in Eq. (2.54)). Earlier we discussed that the second-order response function along the white dashed diagonals in Fig. 2.1 quantifies the higher harmonics and that values of the second-order response function off the white dashed diagonals quantify the mixed response. For the subthreshold regime, the amplitude of the second-order response function shows a pronounced peak on the white dashed diagonal in Fig. 2.1a, but only weak contributions off the white dashed diagonals, indicating a weak mixed response. In the suprathreshold regime, the amplitude of the second-order response function exhibits not only peaks on the white dashed diagonals but also pronounced peaks and stripes off the white diagonals, indicating a strong mixed response. Therefore, in the following we focus on the suprathreshold firing regime. The two cosines which together comprise the full signal are shown in Fig. 2.7a,b. The frequencies of the two cosines satisfy $(\omega_1 + \omega_2)/2\pi \approx r_0$. The nonlinear response to the full signal is shown in Fig. 2.7e and is markedly different from the prediction from the linear theory (given by the zeroth- and first-order terms in Eq. (2.55)).

In order to further illustrate the significance of the mixed response, we also stimulated the LIF model by each of the cosines separately. The responses of the LIF model to the single cosines are shown in Fig. 2.7c,d. Note that, the responses to the single cosines are very well described by the linear theory, indicating that the higher harmonics and the second-order corrections to the time-averaged firing rates are negligible for our choice of signal frequencies and model parameters. By summing the responses to the single cosines and subtracting the unperturbed firing rate, r_0, we obtain the grey curve in Fig. 2.7. The difference between the grey curve and the full nonlinear response (red curve) in Fig. 2.7e is completely given by the mixed-response terms in Eq. (2.55) which are enclosed in $[...]_{MR}$. It is striking, that the mixed response alters the firing rate both with respect to the overall amplitude as well as to the timing of local minima and maxima of the rate.

2.3 Emergence of the weakly nonlinear response for changing signal amplitude, signal frequency and noise strength

The weakly nonlinear theory in Eq. (2.1) was obtained by performing a Volterra expansion of the time-dependent firing rate with respect to the signal and by neglecting terms which are of higher order than ε^2. The nature of this expansion raises two questions. First, for which values of ε do the second-order terms significantly improve the linear response? Second, for which values of ε does weakly nonlinear theory fail to predict the time-dependent firing rate because of the neglected higher-order terms in ε? In order to answer these two questions, we consider the squared relative error between the

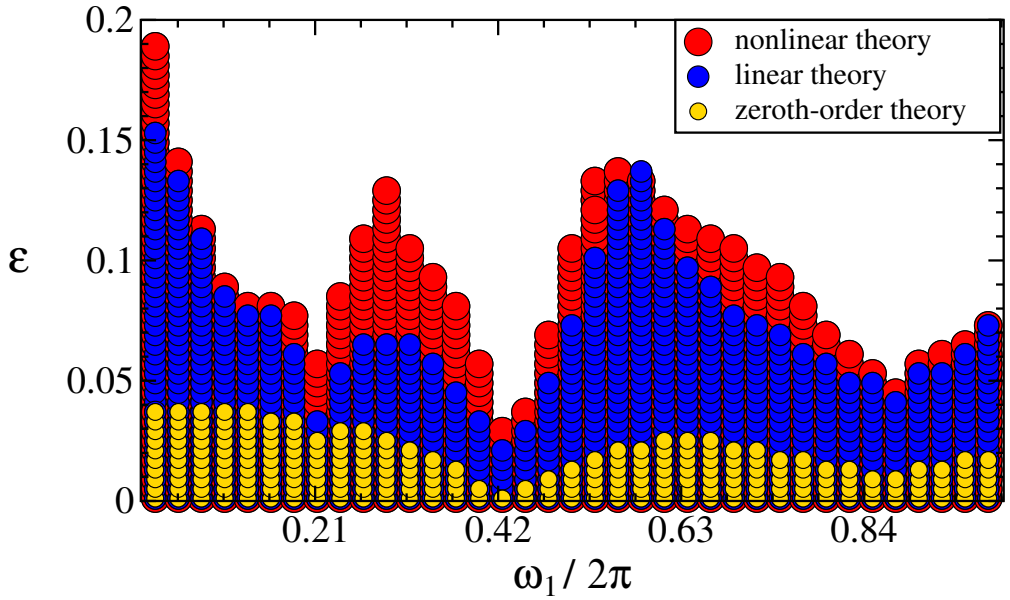

Figure 2.8: For signal frequencies $r_0/2 \leq \omega_1/2\pi \leq r_0 \approx 0.42$ the weakly nonlinear theory outperforms the linear theory for a substantial range of signal amplitudes. Red dots illustrate for which signal frequencies (indicated on the x axis) and for which signal amplitudes (indicated on the y axis) the weakly nonlinear theory, Eq. (2.55), correctly predicts the numerically measured LIF firing rate. Blue dots illustrate the correct prediction of the LIF firing rate by the linear theory (obtained by omitting terms which are proportional to ε^2 in Eq. (2.55)). Yellow dots illustrate that the LIF firing is correctly predicted by the unperturbed firing rate (obtained by setting $\varepsilon = 0$ in Eq. (2.55)). The prediction of a theory is evaluated as correct if the time-averaged mean squared error, Eq. (2.70), does not exceed 10% for the respective theory. Model parameters are as in the suprathreshold regime in Fig. 2.1c,d.

time-dependent firing rate and the prediction from the weakly nonlinear theory

$$\Delta r^2(t) = \left(\frac{r_{\text{sim}}(t) - r_{\text{theo}}(t)}{r_{\text{sim}}(t)} \right)^2 , \tag{2.69}$$

where $r_{\text{sim}}(t)$ is the numerically measured firing rate for the LIF model. The theoretical prediction of the firing rate, $r_{\text{theo}}(t)$, in Eq. (2.69) is either obtained by considering all terms in Eq. (2.55) (weakly nonlinear theory), only zeroth-order and first-order terms in ε in Eq. (2.55) (linear theory) or only the ε-independent term in Eq. (2.55) (zeroth-order theory). In order to quantify the deviation of the weakly nonlinear theory from the time-dependent firing rate we will consider a signal which consists of a single cosine

($\alpha = 1$, $\beta = 0$, in Eq. (2.54)) and consider the time average of Eq. (2.69),

$$\Delta r_T^2 = \frac{1}{T} \int_0^T dt\ \Delta r^2(t)\ , \tag{2.70}$$

where we integrate over one signal period ($T = 2\pi/\omega_1$). In Fig. 2.8 we show for which signal frequencies and which signal amplitudes the weakly nonlinear theory (red circles), the linear theory (blue circles), and the zeroth-order theory (yellow circles) correctly predict the numerically measured time-dependent firing rate of the LIF model. The prediction of a theory is evaluated as correct if the time-averaged mean squared error, Eq. (2.70), does not exceed 10% for the respective theory.

For very small signal amplitudes, ε, the LIF firing rate is correctly described by the zeroth-order theory. The value of ε at which Δr_T^2 for the zeroth-order theory exceeds 10% depends on the signal frequency. This is so because the strength of the modulation of the firing rate depends on the signal frequency. For example, for frequencies $\omega_1/2\pi < 0.15$ the modulations of the LIF rate are so small that even for $\varepsilon = 0.04$ the numerically measured rate can be well described by the zeroth-order theory. The strongest modulations of the firing rate are observed for signal frequencies which are close to the unperturbed firing rate of the LIF model ($r_0 \approx 0.42$). For these frequencies, the zeroth-order theory fails to correctly describe the LIF firing rate even for $\varepsilon = 0.005$.

The range of the ε-values for which the linear theory is correct also depends on the signal frequency. For $\omega_1/2\pi \approx 0.42$ and $\omega_1/2\pi \approx 0.21$ the linear theory already fails for $\varepsilon = 0.004$ and $\varepsilon = 0.04$, respectively. At these frequencies, the LIF firing rate is highly sensitive to the signal and responds in a nonlinear fashion. The linear theory fails in this case, because higher-order contributions of the signal to the modulation of the time-dependent firing rate are ignored. For slow signals ($\omega_1/2\pi < 0.3$) or signals which are faster than the firing rate ($\omega_1/2\pi > 0.5$) the linear theory correctly predicts the LIF firing rate for signal amplitudes up to $\varepsilon = 0.15$ and $\varepsilon = 0.14$, respectively. Note, that the linear theory is also correct for small values of ε and even for $\varepsilon = 0$ but that the blue circles are hidden behind the yellow circles in Fig. 2.8.

The weakly nonlinear theory, correctly describes the LIF firing rate for a wider range of ε-values than the linear theory irrespective of the signal-frequency. For signal frequencies $r_0/2 \leq \omega_1/2\pi < r_0$ the weakly nonlinear theory works particularly well and correctly describes the time-dependent LIF firing rate for signal amplitudes which are much stronger than the amplitudes for which the linear theory works. For $\omega_1/2\pi \approx 0.29$ the weakly nonlinear theory is correct for signal amplitudes up to $\varepsilon = 0.13$ whereas the linear theory already fails at $\varepsilon = 0.07$. For $\omega_1/2\pi \approx 0.42$ the weakly nonlinear theory fails to predict the LIF firing rate for small values of ε. Apparently, the higher-order terms which we neglected in the expansion in Eq. (2.1) are very dominant for signal frequencies which are close the unperturbed firing rate of the LIF model. Note again, that the weakly nonlinear theory also works for small values of ε and even for $\varepsilon = 0$ but that the red circles are hidden behind the blue circles and the yellow circles in Fig. 2.8.

Let us now illustrate how the time-dependent firing rate transitions from the linear

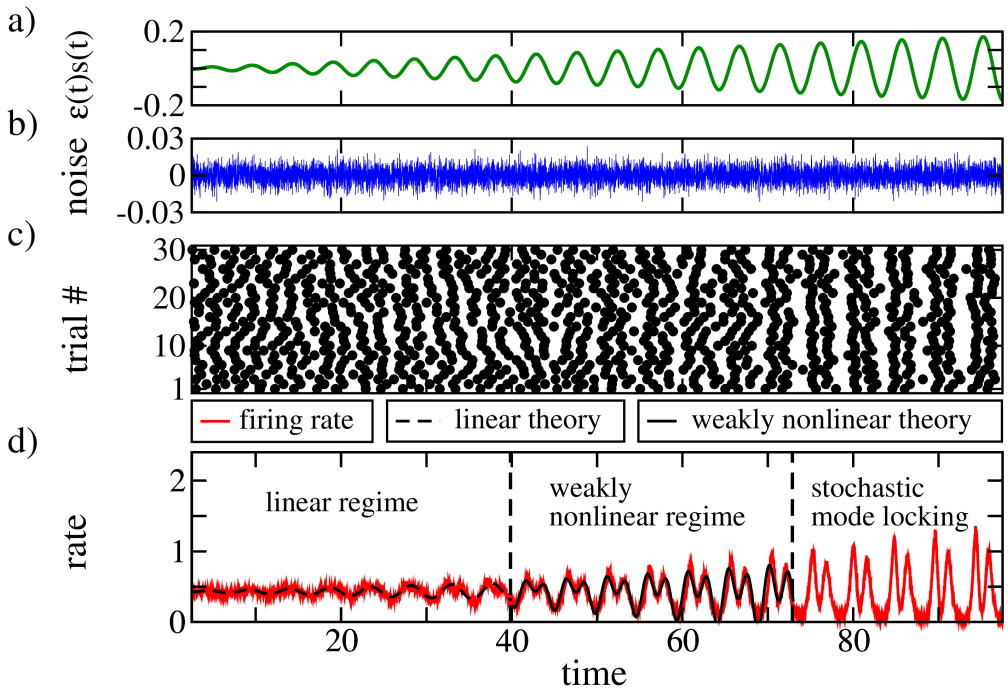

Figure 2.9: The weakly nonlinear response emerges as a transition between the linear response and stochastic mode locking for an increasing signal amplitude. a) Cosine signal with $\omega_1/(2\pi) = 0.21$ multiplied by a time-dependent amplitude $0 < \varepsilon(t) < 0.18$; b) filtered version of the intrinsic noise with $D = 0.002$; c) raster plot; d) firing rate (red) with the linear theory (dashed black line) for $0 < t < 40$ and the weakly nonlinear theory (solid black line) for $40 \leq t \leq 73$; The first and second vertical dashed lines indicate at which point in time the relative mean squared error between the time-dependent firing rate and the theoretical prediction exceeds 10% for the linear theory and the weakly nonlinear theory, respectively. Model parameters are as in the suprathreshold regime in Fig. 2.1c,d.

regime into the weakly nonlinear regime for in creasing ε. We consider a single cosine signal ($\alpha = 1$, $\beta = 0$, $\omega_1/(2\pi) = 0.21$ in Eq. (2.54)) with a time-dependent amplitude, $0 < \varepsilon(t) < 0.18$ (see Fig. 2.9a). The firing regime is suprathreshold with $\mu = 1.1$ and the LIF model is subject to a background noise of constant strength, $D = 0.002$ (see Fig. 2.9b). Because a white noise process possesses an infinite variance, we cannot plot $\xi(t)$ directly. Instead we plot in Fig. 2.9b a filtered version of the noise

$$\eta(t) = \int_t^{t+\Delta t} dt' \, \sqrt{2D}\xi(t') \, . \tag{2.71}$$

For small signal amplitudes, $\varepsilon(t) < 0.08$, the firing rate of the LIF model (red curve in Fig. 2.9d) is perfectly described by the linear theory (dashed black line in Fig. 2.9d). For

larger signal amplitudes, the time-dependent squared relative error, Eq. (2.69), exceeds 10%, which is our criterion for failure of the linear theory. For $0.08 \leq \varepsilon(t) \leq 0.15$ the firing rate is accurately described by the weakly nonlinear theory from Eq. (2.55) (solid black line in Fig. 2.9d). Interestingly, the range over which the weakly nonlinear theory accurately describes the firing rate and at the same time significantly outperforms the linear theory is similar to the ε-range over which the linear theory is valid. The weakly nonlinear theory is truncated at the point at which the time-dependent squared relative error between firing rate and theory exceeds 10%, which happens for $\varepsilon \approx 0.15$. For larger signal amplitudes the weakly nonlinear theory starts to predict negative firing rates, which leads to an increase of the mean squared error between the neural firing rate and the theory. For large signal amplitudes ($\varepsilon > 0.15$), the LIF model exhibits 2:1 mode locking behavior [88] in the sense that it always fires two spikes within one signal period, which can be seen from the raster plot in Fig. 2.9c. In this regime the signal is so strong, that it almost deterministically drives the voltage to the threshold and the intrinsic noise shifts the spikes to earlier or later times. However, the noise can neither prevent the model from spiking nor can it induce additional spikes. Due to the pronounced periods of silence, which can be seen from Fig. 2.9d, and the periods in which the firing rate is very high it seems unlikely that this regime can be described by a perturbation approach in which the time-dependent response is treated as a weak modulation of the unperturbed firing rate. Therefore, it is not surprising that in the regime in which stochastic mode locking is observed our weakly nonlinear theory fails. Instead, the weakly nonlinear theory seems to describe an intermediate regime between linear response and stochastic mode locking.

A somewhat surprising finding is that the weakly nonlinear regime not only emerges for increasing signal amplitude but also for decreasing strength of the noise. As in Fig. 2.9, the signal is again a single cosine ($\alpha = 1$, $\beta = 0$, $\omega_1/(2\pi) = 0.21$ in Eq. (2.54)) but the signal amplitude is kept constant at $\varepsilon = 0.18$ (see Fig. 2.10a). The noise strength is now time-dependent and decreases linearly in time from $D_{max} = 0.014$ to $D_{min} = 0.001$. As before, we cannot plot the white noise process directly, but we show in Fig. 2.10b the filtered noise $\eta(t)$ as in Eq. (2.71). The linear theory and the weakly nonlinear theory are truncated at a point at which the time-dependent mean squared error between neural firing rate and the respective theory exceeds 10%. Additionally, the weakly nonlinear theory is only plotted for noise values for which it outperforms the linear theory. For $t < 30$ the firing rate (red curve in Fig. 2.10d) is well described by the linear theory (dashed black line in Fig. 2.10d). For $t > 63$, the LIF exhibits mode locking behavior for which neither the linear theory nor the weakly nonlinear theory can describe the firing rate correctly. For $30 \leq t \leq 63$ the weakly nonlinear theory emerges as a transition between linear response and stochastic mode locking. The transition from the linear regime to a nonlinear regime due to decreasing noise is consistent with the fact that noise is known to linearize signal-processing systems [89].

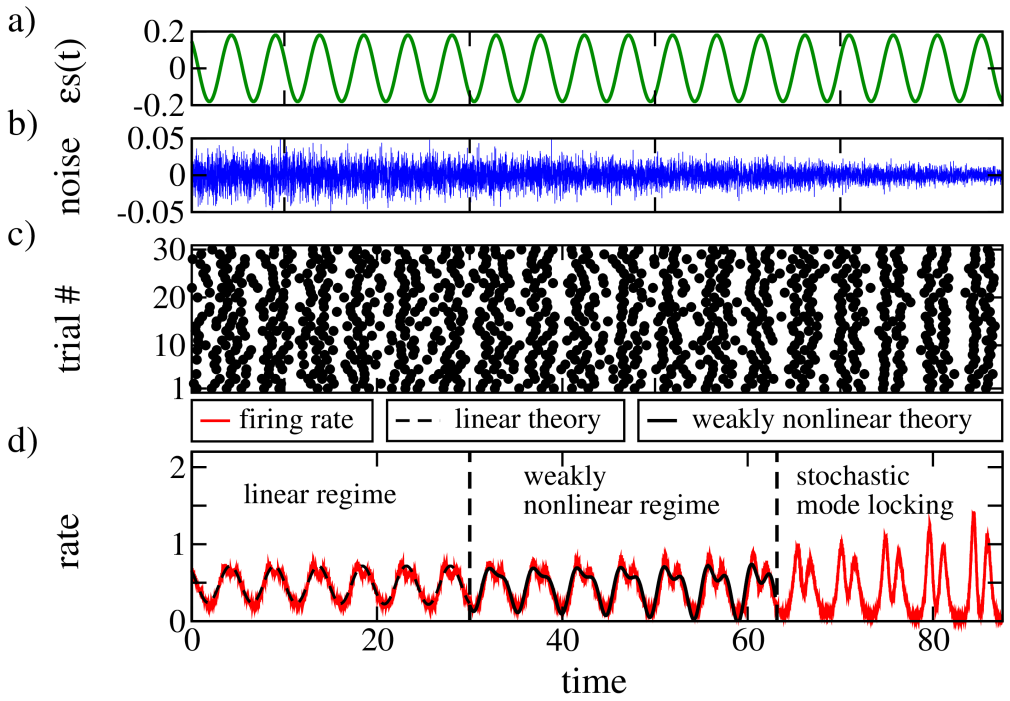

Figure 2.10: The weakly nonlinear response emerges as a transition between the linear response and stochastic mode locking for decreasing strength of the intrinsic noise. a) Cosine signal with $\omega_1/(2\pi) = 0.21$ multiplied by $\varepsilon = 0.18$; b) filtered version of the intrinsic noise with $0.014 > D(t) > 0.001$; c) raster plot; d) firing rate (red) with the linear theory (dashed black line) for $0 < t < 30$ and the weakly nonlinear theory (solid black line) for $30 \leq t \leq 63$; The first and second vertical dashed lines indicate at which point in time the relative mean squared error between the time-dependent firing rate and the theoretical prediction exceeds 10% for the linear theory and the weakly nonlinear theory, respectively. Model parameters are as in the suprathreshold regime in Fig. 2.1c,d.

2.4 Estimation of the first- and second-order response functions from experimental or numerical data

So far we have considered the LIF model, for which the first and second-order response functions were derived analytically. However, it is also possible to estimate χ_1 and χ_2 from neural spike trains which are either obtained experimentally, or via numerical simulations. To this end, it is first necessary to determine the unperturbed firing rate, i.e. the firing rate of the neuron in absence of a stimulus, which for sufficiently long time-windows can be obtained via

$$r_0 = \frac{\langle N_0 \rangle}{T} \,, \tag{2.72}$$

where $\langle N_0 \rangle$ is the spike count of the unperturbed neuron within a time window of length T averaged over different trials. Once the unperturbed firing rate is computed, the neuron is then repeatedly stimulated by the input current $I_1(t) = \varepsilon s_1(t)$ with

$$s_1(t) = \cos(\omega_1 t) \tag{2.73}$$

and the resulting spike trains are recorded. The time-dependent firing rate, $r(t)$, is then determined from the neural spike trains by dividing the fraction of trials in which the neuron has fired in the interval $[t, t + \Delta t]$ by the size of the interval, Δt. Finally, the Fourier modes of $r(t)$ can be related to the functions $\chi_1(\omega_1)$ and $\chi_2(\omega_1, \omega_1)$. In order to obtain the second-order response function for all possible combinations of frequency arguments, the stimulation of the neuron has to be repeated, this time with the input current $I_2(t) = \varepsilon s_2(t)$, where

$$s_2(t) = [\cos(\omega_1 t) + \cos(\omega_2 t)] \tag{2.74}$$

with $\omega_1 > \omega_2$. Now we want to relate the Fourier modes of the neural response with the functions χ_1 and χ_2. For this purpose, we first consider the Fourier transform of the time-dependent firing rate which is observed within a time-window [0,T]

$$\tilde{r}_T(\omega) = \int_0^T dt' \; e^{i\omega t'} r(t) \; . \tag{2.75}$$

Inserting the Volterra expansion for the time-dependent firing rate, Eq. (1.10), into Eq. (2.75) and considering only strictly positive frequencies, $\omega > 0$, we obtain

$$\begin{aligned} \tilde{r}_T(\omega) =& \varepsilon \int_0^T dt' \; e^{i\omega t'} \int_{-\infty}^{t'} dt_1' \; K_1(t' - t_1') s(t_1') \\ &+ \varepsilon^2 \int_0^T dt' \; e^{i\omega t'} \int_{-\infty}^{t'} dt_2' \int_{-\infty}^{t'} dt_2'' \; K_2(t' - t_2', t' - t_2'') s(t_2') s(t_2'') \, . \end{aligned} \tag{2.76}$$

Strictly speaking, the lower integration boundaries of the convolution integrals in the above equation should be adjusted to zero, because per definition the signal is switched on at $t = 0$ and no previous values of the signal can affect the response at time t. In practice, however, for sufficiently large time windows, T, the kernels K_1 and K_2 will decay sufficiently fast, such that the lower integration boundaries of the convolution integrals will not affect the response and can safely be set to minus infinity. Intuitively, this corresponds to the reasonable assumption that after a short transient period the response at time t is not influenced by the signal at the beginning of the experiment.

Via a change of variables the first term in Eq. (2.76) can be rewritten into

$$\int_0^T dt'\, e^{i\omega t'} \int_{-\infty}^{t'} dt_1'\, K_1(t'-t_1')s(t_1') = \int_0^T dt'\, e^{i\omega t'} \int_{-\infty}^{\infty} dt_1'\, K_1(t_1')s(t'-t_1')\,, \tag{2.77}$$

where we exploited the fact that the linear filter K_1 is a causal filter with the property $K_1(t) = 0$ for $t < 0$ allowing us to set the lower integration boundary of the convolution integral to minus infinity. Similarly, we can also rewrite the second term in Eq. (2.76) into

$$\begin{aligned}\int_0^T dt'\, e^{i\omega t'} \int_{-\infty}^{t'} dt_2' \int_{-\infty}^{t'} dt_2''\, K_2(t'-t_2', t'-t_2'')s(t_2')s(t_2'') = \\ = \int_0^T dt'\, e^{i\omega t'} \int_{-\infty}^{\infty} dt_2' \int_{-\infty}^{\infty} dt_2''\, K_2(t_2', t_2'')s(t'-t_2')s(t'-t_2'')\,.\end{aligned} \tag{2.78}$$

Now we want to solve Eq. (2.76) with the signal from Eq. (2.73). Therefore, we first note that the signal in Eq. (2.73) can be rewritten in terms of complex exponentials

$$s_1(t) = \frac{1}{2}\left(e^{i\omega_1 t} + e^{-i\omega_1 t}\right)\,. \tag{2.79}$$

Inserting Eq. (2.79) into Eq. (2.77), employing the definition of the function χ_1 from Eq. (2.2) and again only considering positive frequency arguments, $\omega > 0$, we find

$$\varepsilon \int_0^T dt'\, e^{i\omega t'} \int_{-\infty}^{t'} dt_1'\, K_1(t'-t_1')s(t_1') = \begin{cases} \frac{\varepsilon T}{2}\chi_1(\omega_1) & \text{for} \quad \omega = \omega_1\,, \\ 0 & \text{else}\,. \end{cases} \tag{2.80}$$

Using Eq. (2.3) we can also simplify Eq. (2.78) into

$$\varepsilon^2 \int_0^T dt'\, e^{i\omega t'} \int_{-\infty}^{t'} dt_2' \int_{-\infty}^{t'} dt_2''\, K_2(t'-t_2', t'-t_2'')s(t_2')s(t_2'') = \begin{cases} \frac{\varepsilon^2 T}{4}\chi_2(\omega_1,\omega_1) & \text{for} \quad \omega = 2\omega_1\,, \\ 0 & \text{else}\,. \end{cases} \tag{2.81}$$

Employing Eq. (2.76) and Eq. (2.80) we find that the first-order response function is given by

$$\chi_1(\omega_1) = \frac{2}{\varepsilon T}\tilde{r}_T(\omega_1)\,, \tag{2.82}$$

where ω_1 is the angular frequency of the cosine signal in Eq. (2.73). Analogously we find from Eq. (2.76) and Eq. (2.81) that

$$\chi_2(\omega_1, \omega_1) = \frac{4}{\varepsilon^2 T} \tilde{r}_T(2\omega_1) \; . \tag{2.83}$$

By repeatedly stimulating the neuron with the signal from Eq. (2.73) with different frequencies ω_1 it is possible to measure the full function $\chi_1(\omega)$ and the function $\chi_2(\omega, \omega)$ for equal frequency arguments. The function χ_2 for frequency arguments of opposing sign can be related the mean number of spikes $\langle N \rangle$ via

$$\chi_2(\omega_1, -\omega_1) = \frac{1}{\varepsilon^2} \left(\frac{\langle N \rangle}{T} - r_0 \right) \; . \tag{2.84}$$

Measuring $\langle N \rangle$ for different frequencies ω_1 of the signal $s_1(t)$ allows to determine the full function $\chi_2(\omega, -\omega)$. Note, that in order to obtain $\chi_2(\omega, -\omega)$, it is necessary to know r_0 first. Now, in order to obtain the full function χ_2 we need to stimulate the neuron with the signal $s_2(t)$ from Eq. (2.74). Inserting Eq. (2.74) into the Fourier-transformed response, Eq. (2.76), and solving for $\chi_2(\omega_1, \omega_2)$ and $\chi_2(\omega_1, -\omega_2)$ we find

$$\chi_2(\omega_1, \omega_2) = \frac{2}{\varepsilon^2 T} \tilde{r}_T(\omega_1 + \omega_2) \tag{2.85}$$

and

$$\chi_2(\omega_1, -\omega_2) = \begin{cases} \frac{2}{\varepsilon^2 T} \tilde{r}_T(\omega_1 + \omega_2) - \frac{2}{\varepsilon^2 T} \chi_1(\omega_2) & \text{for} \quad \omega_1 = 2\omega_2 \; , \\ \frac{2}{\varepsilon^2 T} \tilde{r}_T(\omega_1 + \omega_2) & \text{else.} \end{cases} \tag{2.86}$$

In order to obtain the full function χ_2 we first choose different frequencies ω_1 for the first cosine in Eq. (2.74). Then, for every ω_1 we also vary the frequency of the second cosine such that $0 < \omega_2 < \omega_1$. Finally, the full function $\chi_2(\omega, \omega')$ is obtained by employing the symmetry relations Eq. (2.51) and Eq. (2.52). From Eq. (2.86) we can also see, that in order to determine the full function χ_2 it is first necessary to know the linear response function χ_1. Therefore, it is not sufficient to stimulate the neuron only with s_2 from Eq. (2.74) but a stimulation with the signal $s_1(t)$ from Eq. (2.73) is also required. Finally, we want to mention, that in the Volterra expansion in Eq. (2.76) we neglected terms of third order in ε and higher. For this reason, the amplitude ε of the input currents $I_1(t)$ and $I_2(t)$ has to be chosen small enough, such that higher order terms in ε do not contribute to the response, but large enough, such that the second-order response can be reliably determined from a finite number of trials.

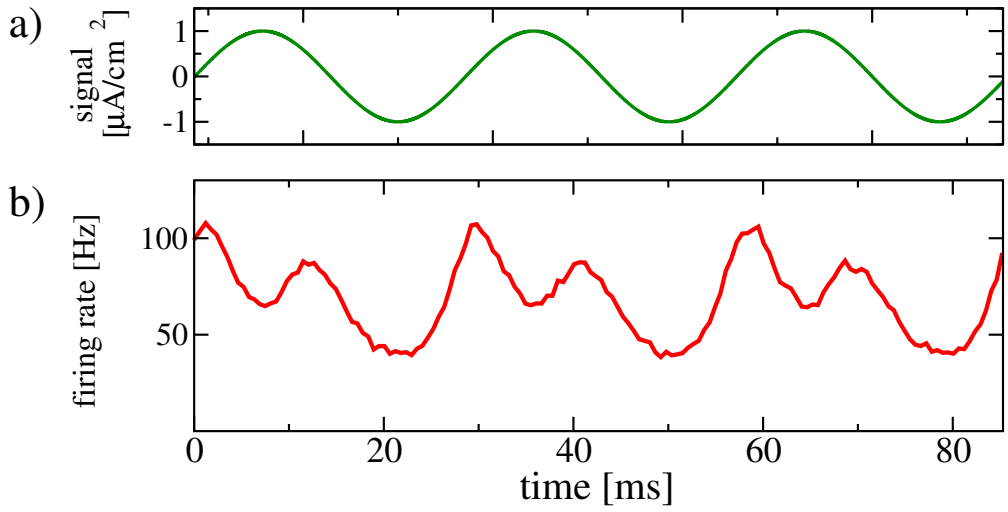

Figure 2.11: In the suprathreshold regime the response of the NaK model to a cosine stimulus exhibits a higher harmonic. a) cosine signal with $f = 35$Hz; b) time-dependent firing rate for $\varepsilon = 0.05$. The mean current was $I_0 = 5\mu\text{A/cm}^2$.

2.5 Weakly nonlinear response of the NaK model with channel noise

The LIF model is a rather simplistic model, for which the spiking mechanism is incorporated via an explicit fire-and-reset rule and for which the intrinsic fluctuations are modeled by additive Gaussian white noise. Here, we verify that the weakly nonlinear response is not a particularity of the LIF model but is also observed for the NaK model, which is endowed with channel noise. In Fig. 2.11 we demonstrate the excitation of a harmonic oscillation in the suprathreshold regime ($I_0 > 4.51\mu\text{A/cm}^2$). The frequency of the input signal, $f = 35$Hz, roughly equals half of the unperturbed firing rate, $r_0 \approx 66$Hz. The harmonic oscillation is very pronounced and seems to be as strong as or even slightly stronger than the ground mode. This behavior is similar to what we already observed for the LIF model in Fig. 2.3. We also employ the numerical routine for the estimation of the second-order response function which we outlined in the previous section. The numerically measured second-order response function (Fig. 2.12) shows a striking resemblance with the second-order response function which we calculated analytically for the LIF model (Fig. 2.1). In the subthreshold regime the amplitude of the second-order response function exhibits two peaks, which are located at $\omega_1/2\pi = \omega_2/2\pi = \pm 30$Hz. This shape indicates, that the strongest weakly nonlinear contribution to the response stems from the higher harmonic. In the suprathreshold regime the amplitude of the second-order response function exhibits triangular structures, which indicates that not only the higher harmonic but also the mixed response contributes to nonlinear rate modulations. Also the overall dependence of the phase on the two frequency arguments are in both firing regimes comparable to what we found for the LIF model in Fig. 2.1.

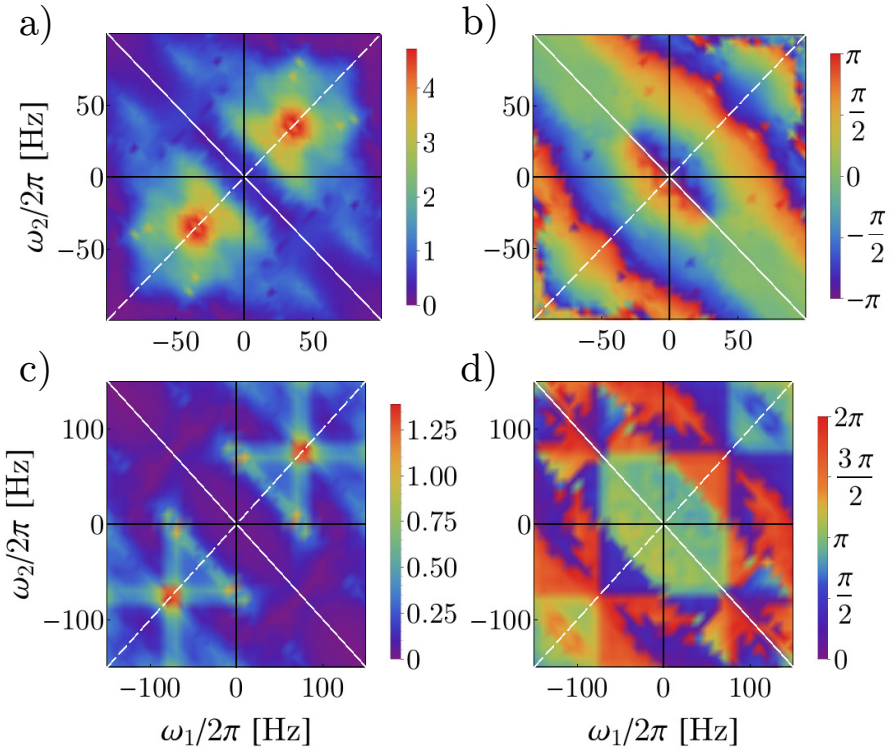

Figure 2.12: Amplitude (a,c) and phase (b,d) of the second-order response function χ_2 for the NaK model in the subthreshold (a,b) and suprathreshold (c,d) regimes. $I_0 = 4.4\mu\text{A/cm}^2$ in the subthreshold regime and $I_0 = 5\mu\text{A/cm}^2$ in the suprathreshold regime. For the numerical estimates of the response function we employed input currents with $\varepsilon = 0.05$ in the subthreshold regime and $\varepsilon = 0.1$ in the suprathreshold regime. The amplitude in a) and c) is given in $\text{kHz}^2\text{cm}^4/\mu\text{A}^2$.

We also find some differences between the NaK and the LIF model. First, for the NaK model the amplitude of the second-order response function is higher in the subthreshold than in the suprathreshold regime, while the opposite behavior was observed for the LIF model. Second, the detailed behavior of the phase in Fig. 2.12d, which is almost constant for small frequency arguments, differs from the behavior of the phase in Fig. 2.1d, in which case it changes by π in the same frequency range (in relation to the firing rate). Third, we find that in the suprathreshold regime for the NaK model, the peak of the amplitude of χ_2 on the dashed white diagonal in Fig. 2.1c is significantly larger than the peaks off the diagonal, which was not the case for the LIF model. The last difference could potentially be explained by the numerical estimation of the second-order response for the NaK model as compared to the exact analytic result for the LIF model. If the amplitude ε of the stimulation which is employed to estimate χ_2 is too large, this could lead to an underestimation of the peaks of χ_2, because higher-order contributions in ε will in general decrease the effect of the input on the rate. Another reason for why the peaks

in Fig. 2.12c are not as pronounced as in Fig. 2.1c could be due to the coarse frequency resolution of χ_2 for the NaK model. If the frequency resolution is not sufficiently small, then very narrow large peaks could potentially be missed. For the computationally costly NaK model, we were not able to reduce ε or the frequency resolution much more than what we have used here. Despite some differences however, the qualitative similarities of χ_2 in the two regimes of the LIF and NaK models imply that the weakly nonlinear response is a regime which is not only relevant for the LIF model but could be of a more general relevance.

2.6 Summary

At the beginning of this chapter we expressed the time-dependent firing rate by a Volterra expansion with respect to a time-dependent signal which was truncated after the second-order. For the LIF model we calculated analytically the second-order response function, which quantifies second-order contributions of the input signal to modulations of the firing rate. In terms of the response to cosine signals, we provided an intuitive interpretation of the second-order response function, which is not only applicable to the LIF model but can serve as a general interpretation of the second-order response functions of neuron models or real neurons. For the LIF model we then demonstrated that in firing regimes which are dominated by the mean input and where the intrinsic noise is weak the second order response can be very pronounced. In such a weakly nonlinear regime we demonstrated for the LIF model the excitation of higher harmonic oscillations in the firing rate by a cosine signal, a nonlinear increase or decrease of the time-averaged firing rate, and the excitation of mixed harmonics by a sum of cosine signals. The weakly nonlinear theory can not only be applied to cosine signals but also to non-oscillatory signals with a general time dependence. For a Gaussian signal with a power spectrum which is peaked around half of the natural frequency of the neuron, we demonstrated how the weakly nonlinear theory outperforms the linear response. Interestingly, the emergence of the weakly nonlinear regime is not only determined by the signal amplitude, as suggested by the nature of the Volterra expansion for the firing rate, but is also dependent on the frequency of the input signal and on the level of the intrinsic noise of the neuron. Finally, we discussed how the second-order response function can be determined from experimental data and demonstrated the numerical procedure on the NaK model which is endowed with a discrete channel noise. Similar to the LIF model, the NaK exhibited a pronounced weakly nonlinear response, which suggests that our results for the weakly nonlinear response are not specific to the LIF model but are of a more general nature.

3 Nonlinear decoding of static signals from the neural spike count

Abstract | This chapter is devoted to the extension of the linear lower bound for the mutual information between an input signal and the neural output. For the special case of a static input signal we derive a quadratic lower bound, which quantifies the minimal information that is recovered from the neural output by means of an optimal quadratic filter and also derive an upper bound for the mutual information. For the LIF neuron model in the mean-driven regime we show that a quadratic reconstruction of the signal can recover significantly more information from the output than an optimal linear reconstruction. Furthermore, we find that the quadratic lower bound provides a correct estimate for the exact mutual information for a larger range of signal strengths than the linear lower bound.

In chapter 2 we studied the dependence of the neural output on higher moments of the signal. Here, we take a different approach and consider the reconstruction of the input signal from higher moments of the output. A linear reconstruction of a time-dependent input signal, $s(t)$, from the spike train, $x(t)$, is in general given by [38]

$$s_{rec}^{lin}(t) = \int_{-\infty}^{\infty} d\tau \; h(\tau) x(t-\tau) \; , \tag{3.1}$$

where $h(\tau)$ is an acausal linear filter. The above equation can be extended to also include second-order moments of the output

$$s_{rec}^{quad}(t) = \int_{-\infty}^{\infty} d\tau \; h'(\tau) x(t-\tau) + \int_{-\infty}^{\infty} d\tau_1 \int_{-\infty}^{\infty} d\tau_2 \; g(\tau_1,\tau_2) x(t-\tau_1) x(t-\tau_2) \; , \tag{3.2}$$

where $h'(\tau)$ and $g(\tau_1,\tau_2)$ are the linear and nonlinear filters of the second-order reconstruction, respectively. For a static input signal, the time dependence of the spike train does not carry any information about the signal. In this case, all the relevant information about the signal s is contained in the spike count N and Eq. (3.1) and Eq. (3.2)

simplify to

$$s_{rec}^{lin} = h \cdot N \quad \text{and} \quad s_{rec}^{quad} = h' \cdot N + g \cdot N^2 \; . \tag{3.3}$$

One possible way to asses the quality of the linear and the quadratic reconstructions is to define a reconstruction error $\eta = s - s_{rec}$ and to choose h, h' and g such that the variance σ_η^2 is minimized. If the quadratic reconstruction allowed to achieve a much smaller error than the linear reconstruction, one could safely say that the decoding of the stimulus (and, consequently, also the encoding) is nonlinear. The drawback of this method is that there is no way to asses how much information is actually contained in the output and whether the reconstructions in Eq. (3.3) come close to extracting the full information or not. Therefore, in the next section, we will derive lower bounds of the mutual information, which will allow to quantify the minimal information that is extracted from the output by an optimal linear or quadratic filter and will also allow to compare the retrieved information with the total information that is contained in the output.

3.1 Derivation of the quadratic lower bound for the mutual information

In this chapter we consider a Gaussian signal variable s that is processed by an unspecified noisy nonlinear system that generates an output x. The amount of information that is transmitted by the system is quantified by the mutual information $MI_{x,s}$, Eq. (1.27). Here, we will derive the well-known lower bound for the mutual information [38] that does not require the knowledge of the conditional distribution $p(x|s)$ or the joint distribution $p(x, s)$ but can be computed from simple statistics of input s and output x. Although the linear lower bound can also be computed for time-dependent processes [38], in this chapter we focus on stochastic variables only. This will allow us to extend the derivation of the linear lower bound and derive an improved quadratic lower bound.

In a first step we consider a reconstructed signal s_{rec} that is obtained from a deterministic transformation of the output x. Since a deterministic transformation does not incorporate any knowledge about the signal except the one that is already contained in x it follows that

$$MI_{s,x} \geq MI_{s,s_{rec}} \; , \tag{3.4}$$

which is also known as *data processing inequality* [90]. Following the definition in Eq. (1.25), the mutual information between s and s_{rec} can be rewritten in terms of entropies as

$$MI_{s,s_{rec}} = H_s - \left\langle H_{s|s_{rec}} \right\rangle_{s_{rec}} \; , \tag{3.5}$$

where H_s is the signal entropy and $H_{s|s_{rec}}$ is the noise entropy for a fixed realization of

s_{rec}. The noise in this case is not the intrinsic noise of the signal processing system but a reconstruction noise that is defined as

$$\eta(s_{rec}) = s - s_{rec} \ . \tag{3.6}$$

Intuitively, for a given reconstructed signal s_{rec}, the variable η represents the uncertainty about the input signal s that arises due to the nonlinearity of the signal processing system, due to the intrinsic noise of the system and due to the (potentially) suboptimal choice of the deterministic reconstruction. Because for a known value of s_{rec}, all of the uncertainty about s is captured by η, we can write for the conditioned noise entropy in Eq. (3.5)

$$H_{s|s_{rec}} = H_{\eta|s_{rec}} \ . \tag{3.7}$$

In general, the reconstruction noise $\eta(s_{rec})$ and its statistics will depend on the specific realization of s_{rec}. In particular, also the variance of the reconstruction noise, $\sigma_\eta^2(s_{rec})$, will in general be a function of s_{rec}. Furthermore, the reconstruction noise is in general not Gaussian distributed, which means that in general it is not possible to find a simple analytical expression for $H_{\eta|s_{rec}}$. However, it is possible to find a simple estimate of the conditioned noise entropy by considering a Gaussian variable η_G with variance $\sigma_\eta^2(s_{rec})$. Utilizing the fact that for a given variance, the entropy is maximized by a Gaussian random variable η_G [57], we obtain an upper bound for the entropy of the reconstruction noise:

$$H_{\eta|s_{rec}} \leq H_{\eta_G|s_{rec}} \ . \tag{3.8}$$

Taking further into account that an averaged conditional entropy is always smaller than the unconditioned entropy [38] leads to

$$H_{\eta_G|s_{rec}} \leq H_{\eta_G} \ , \tag{3.9}$$

where H_{η_G} is now the entropy of a Gaussian variable with the variance σ_η^2 which is not conditioned on s_{rec} anymore. In conclusion, using Eq. (3.4), Eq. (3.5), Eq. (3.7), Eq. (3.8) and Eq. (3.9) we find that

$$MI_{s,x} \geq H_s - H_{\eta_G} \ , \tag{3.10}$$

where H_s and H_{η_G} are both differential entropies of Gaussian variables s and η_G with variances σ_s^2 and σ_η^2. With the Gaussian densities

$$p_s(s) = \frac{1}{\sqrt{2\pi\sigma_s^2}} e^{-\frac{s^2}{2\sigma_s^2}} \ \text{ and } \ p_\eta(\eta) = \frac{1}{\sqrt{2\pi\sigma_\eta^2}} e^{-\frac{\eta^2}{2\sigma_\eta^2}} \ , \tag{3.11}$$

we can calculate the right side of Eq. (3.10)

$$\begin{aligned} H_s - H_{\eta_G} =& -\int ds\; p_s(s)\log_2\left(p_s(s)\right) + \int d\eta\; p_\eta(\eta)\log_2\left(p_\eta(\eta)\right) \\ =& \log_2\left(\sqrt{2\pi\sigma_s^2}\right) - \int ds\; p_s(s)\frac{s^2}{2\sigma_s^2} - \log_2\left(\sqrt{2\pi\sigma_\eta^2}\right) + \int d\eta\; p_\eta(\eta)\frac{\eta^2}{2\sigma_\eta^2} \\ =& -\frac{1}{2}\log_2\left(\frac{\sigma_\eta^2}{\sigma_s^2}\right)\ . \end{aligned} \tag{3.12}$$

Combining Eq. (3.12) and Eq. (3.10) it follows that

$$MI_{s,x} \geq -\frac{1}{2}\log_2\left(\frac{\sigma_\eta^2}{\sigma_s^2}\right)\ . \tag{3.13}$$

The above relation is true for any signal processing system. The equality in the above relation holds if the deterministic signal reconstruction extracts all the relevant information from the output, i.e. if the reconstruction is optimally chosen, and if the statistics of the reconstruction noise are Gaussian and independent of the realization of s_{rec} (see Eq. (3.8) and Eq. (3.9)). On the other hand, if the reconstruction noise is non-Gaussian, if its statistics are not independent of the realization of s_{rec} and if the deterministic reconstruction does not extract all the information about the signal that is contained in the output, the lower bound will always be lower than the true mutual information and could even be equal to zero.

Linear reconstruction

We now want to consider a linear reconstruction of the signal,

$$s_{rec} = h \cdot x\ , \tag{3.14}$$

where h is a constant coefficient. The variance of the reconstruction noise is computed by combining Eq. (3.6) and Eq. (3.14)

$$\begin{aligned} \sigma_\eta^2 =& \Big\langle \eta^2 \Big\rangle - \Big\langle \eta \Big\rangle^2 \\ =& \Big\langle (s - s_{rec})^2 \Big\rangle - \Big\langle s - s_{rec} \Big\rangle^2 \\ =& \Big\langle s \Big\rangle^2 + h^2 \Big\langle (x - \langle x\rangle)^2 \Big\rangle - 2h \Big\langle s(x - \langle x\rangle) \Big\rangle\ , \end{aligned} \tag{3.15}$$

where we employed that $\langle s\rangle = 0$. The coefficient h is now chosen such that σ_η^2 is minimized, i.e. h is chosen such that the signal reconstruction in Eq. (3.14) is optimal. This is achieved by taking the derivative of the noise variance w.r.t. h and setting it to

zero

$$\frac{\partial}{\partial h}\sigma_{\eta^2} = 0 \,, \tag{3.16}$$

which leads to

$$0 = 2h\Big\langle (x - \langle x\rangle)^2 \Big\rangle - 2\Big\langle s(x - \langle x\rangle)\Big\rangle \,, \tag{3.17}$$

and results for h in

$$h = \frac{\langle s(x - \langle x\rangle)\rangle}{\langle (x - \langle x\rangle)^2\rangle} \,. \tag{3.18}$$

It is easy to show that $\frac{\partial^2}{\partial h^2}\sigma_\eta^2 > 0$, implying that the optimal choice of h in Eq. (3.18) indeed minimizes the variance of the reconstruction noise. Inserting Eq. (3.18) into Eq. (3.15) we find

$$\sigma_\eta^2 = \sigma_s^2 \left(1 - \rho_{x,s}^2\right) \,, \tag{3.19}$$

where $\rho_{x,s}^2$ is the squared linear correlation coefficient

$$\rho_{x,s}^2 = \frac{\langle (x - \langle x\rangle)\,(s - \langle s\rangle)\rangle^2}{\sigma_x^2 \sigma_s^2} \,. \tag{3.20}$$

Inserting Eq. (3.19) into Eq. (3.13) we arrive at the final result for the linear lower bound for the mutual information

$$MI_{lb}^{lin} = -\frac{1}{2}\log_2(1 - \rho_{x,s}^2) \,. \tag{3.21}$$

Under the assumption that the reconstruction noise η is Gaussian and independent of the realization of s_{rec}, Eq. (3.21) tells us how much information about the signal can at least be retrieved from the output by a linear reconstruction (by using the optimal linear filter h). In this sense, MI_{lb}^{lin} is a linear measure of information. Note, that Eq. (3.21) only gives the *minimal* information that is retrieved by an optimal linear filter, as we assumed in the derivation, that the reconstruction noise η is Gaussian and independent of the signal, which in general is not necessarily the case. However, for cases in which the above assumptions are fulfilled, the linear lower bound can even serve as an estimator of the exact mutual information. This is particularly useful in cases in which the exact mutual information is not accessible via an analytic or a numerical computation (e.g. due to a limited amount of data).

Quadratic reconstruction

In the previous section we sketched the well-known derivation of the linear lower bound for the mutual information, that takes into account linear correlations between input

and output. Furthermore, the linear lower bound can be interpreted as the minimal information, that is retrieved from the output by an optimal linear filter. In this section we will extend the derivation from the previous section and derive an improved lower bound, that not only takes into account linear statistics of s and x but also statistics of higher orders. In particular, we will consider a second-order reconstruction of the signal from the output and derive the quadratic lower bound, which provides the minimal information that is retrieved from the output by a quadratic filter. By comparison with the linear lower bound, the quadratic lower bound will allow us to identify regimes of neural firing for which the encoding and decoding of input signals are nonlinear.

We start by defining a nonlinear reconstruction of the signal

$$s_{rec} = h' \cdot x + g' \cdot x^2 \,, \tag{3.22}$$

where h' and g' are constant coefficients, that are chosen such that the reconstruction noise is minimized. Note, that because of the additional quadratic term in Eq. (3.22), the coefficient computed in Eq. (3.18) will in general not be optimal anymore. The variance of the reconstruction noise is given by

$$\begin{aligned}
\sigma_\eta^2 =& \Big\langle \eta^2 \Big\rangle - \Big\langle \eta \Big\rangle^2 \\
=& \Big\langle (s - s_{rec})^2 \Big\rangle - \Big\langle s - s_{rec} \Big\rangle^2 \\
=& \Big\langle s^2 \Big\rangle + h'^2 \Big\langle \left(x - \langle x \rangle\right)^2 \Big\rangle + g'^2 \Big\langle \left(x^2 - \Big\langle x^2 \Big\rangle\right)^2 \Big\rangle \\
&+ 2h'g' \Big\langle \left(x - \langle x \rangle\right) \left(x^2 - \Big\langle x^2 \Big\rangle\right) \Big\rangle - 2h' \Big\langle s \left(x - \langle x \rangle\right) \Big\rangle - 2g' \Big\langle s \left(x^2 - \Big\langle x^2 \Big\rangle\right) \Big\rangle \,,
\end{aligned} \tag{3.23}$$

where we used that $\langle s \rangle = 0$. The optimal coefficients h' and g' are determined from the conditions

$$\frac{\partial}{\partial h'} \sigma_\eta^2 = 0 \,, \tag{3.24}$$

$$\frac{\partial}{\partial g'} \sigma_\eta^2 = 0 \,, \tag{3.25}$$

which lead to two linear equations

$$h' = \rho_{s,x} \frac{\sigma_s}{\sigma_x} - g' \cdot \rho_{x,x^2} \frac{\sigma_{x^2}}{\sigma_x} \,, \tag{3.26}$$

$$g' = \rho_{s,x^2} \frac{\sigma_s}{\sigma_{x^2}} - h' \cdot \rho_{x,x^2} \frac{\sigma_x}{\sigma_{x^2}} \,, \tag{3.27}$$

where we defined the variance $\sigma_{x^2} = \sqrt{\Big\langle \left(x^2 - \langle x^2 \rangle\right)^2 \Big\rangle}$ and the squared nonlinear corre-

lation coefficients

$$\rho^2_{x^2,s} = \frac{\left\langle \left(x^2 - \langle x^2 \rangle\right)\left(s - \langle s \rangle\right) \right\rangle^2}{\sigma^2_{x^2}\sigma^2_s} , \tag{3.28}$$

$$\rho^2_{x^2,x} = \frac{\left\langle \left(x^2 - \langle x^2 \rangle\right)\left(x - \langle x \rangle\right) \right\rangle^2}{\sigma^2_{x^2}\sigma^2_x} . \tag{3.29}$$

The two linear equations Eq. (3.26) and Eq. (3.27) can be solved for h' and g', which gives us the optimal reconstruction coefficients

$$h' = \frac{\sigma_s}{\sigma_x} \frac{1}{1-\rho^2_{x,x^2}} \left(\rho_{s,x} - \rho_{s,x^2}\rho_{x,x^2}\right) , \tag{3.30}$$

$$g' = \frac{\sigma_s}{\sigma_{x^2}} \frac{1}{1-\rho^2_{x,x^2}} \left(\rho_{s,x^2} - \rho_{s,x}\rho_{x,x^2}\right) . \tag{3.31}$$

Note, that the optimal linear filter for the quadratic reconstruction h' is different from the optimal filter of the linear reconstruction Eq. (3.18). Inserting h' and g' from Eq. (3.30) and Eq. (3.31) into Eq. (3.23) we find for the variance of the reconstruction noise

$$\begin{aligned} \sigma^2_\eta =& \sigma^2_s - \frac{\sigma^2_s}{1-\rho^2_{x,x^2}} \left(\rho^2_{s,x} + \rho^2_{s,x^2} - 2\rho_{s,x}\rho_{s,x^2}\rho_{x,x^2}\right) \\ =& \sigma^2_s - \sigma^2_s\rho^2_{s,x} - \sigma^2_s \frac{\left(\rho_{s,x^2} - \rho_{s,x}\rho_{x,x^2}\right)^2}{1-\rho^2_{x,x^2}} , \end{aligned} \tag{3.32}$$

Combining Eq. (3.32) and Eq. (3.13) we arrive at the final result for the quadratic lower bound

$$MI^{quad}_{lb} = -\frac{1}{2}\log_2\left(1 - \rho^2_{x,s} - \frac{\left(\rho_{s,x^2} - \rho_{s,x}\rho_{x,x^2}\right)^2}{1-\rho^2_{x,x^2}}\right) . \tag{3.33}$$

Because $\rho^2_{x,x^2} \leq 1$, it holds that

$$\frac{\left(\rho_{s,x^2} - \rho_{s,x}\rho_{x,x^2}\right)^2}{1-\rho^2_{x,x^2}} \geq 0 , \tag{3.34}$$

leading to

$$MI^{quad}_{lb} \geq MI^{lin}_{lb} = -\frac{1}{2}\log_2\left(1-\rho^2_{x,s}\right) . \tag{3.35}$$

The quadratic lower bound corresponds to the minimal information that is recovered

from the output by an optimal quadratic reconstruction and allows to quantify the quality of the quadratic reconstruction. This is particularly useful, because the quality of the optimal quadratic reconstruction, Eq. (3.22), cannot be assessed by one single correlation coefficient, as in the case of the linear lower bound Eq. (3.21). The quadratic lower bound, Eq. (3.33), combines the contributions of different nonlinear correlation coefficients into one measure of the quality of information transmission. By comparing the linear and the quadratic lower bound to the exact mutual information between input and output it is not only possible to compare the linear and quadratic reconstructions with each other but also to asses whether they recover all of the information that is present in the output or not. Furthermore, for weak signal amplitudes and weakly nonlinear systems for which the linear and second-order filters are sufficient to reconstruct the signal, the linear or quadratic lower bounds could be used as approximations for the exact mutual information.

3.2 Derivation of an upper bound for the mutual information for systems with signal-dependent Gaussian noise

In this section we will derive an upper bound for the mutual information between a static signal and an output variable the variability of which is signal dependent and Gaussian distributed. The derivation follows the arguments from the references [38, p. 126] and [62], in which the authors consider time-dependent input signals and time-dependent output variables. In both references, the authors discuss an upper bound for the mutual information for which the output variability for a fixed signal is Gaussian and does not exhibit any signal dependence. Here, we consider information transmission for static input and output variables but for systems the output of which exhibits a Gaussian distributed variability with an explicit signal dependence. In particular, we consider the output variable

$$x = M(s) + \sqrt{V(s)}\,\zeta\ , \tag{3.36}$$

where the functions $M(s)$ and $V(s)$ are arbitrary (nonlinear) functions of the input signal s and the variable ζ is a Gaussian noise with unit variance and zero mean. As before, the key quantity of interest is the mutual information (cf. Eq. (1.25)), which quantifies the average amount of information which is conveyed by x about the signal s. However, in contrast to the previous section, we will not assume that the distribution of the signal s is Gaussian.

For a fixed signal, all the variability of x in Eq. (3.36) comes from the noise, which is multiplied with $\sqrt{V(s)}$. Consequently, the density of x conditioned on a realization of the signal reads

$$p(x|s) = \frac{1}{\sqrt{2\pi V(s)}} \exp\left[-\frac{(x - M(s))^2}{2V(s)}\right]\ . \tag{3.37}$$

For the differential noise entropy of the output x, the above equation leads to

$$\begin{aligned}\left\langle H_{x|s}\right\rangle_s =& -\int ds\; p(s)\int dx\; p(x|s)\log_2(p(x|s))\\ =&\frac{1}{2}\int ds\; p(s)\log_2\left(2\pi V(s)\right)-\frac{1}{2}\\ =&\frac{1}{2}\left\langle \log_2\left(2\pi V(s)\right)\right\rangle_s-\frac{1}{2}\;.\end{aligned} \tag{3.38}$$

In order to compute the mutual information between x and s we now need to find a simple expression for the output entropy H_x. In general, the output x in Eq. (3.36) will be non-Gaussian and for non-Gaussian variables only few cases exist in which a simple analytical expression for H_x can be found. So, instead of the non-Gaussian variable x with variance σ_x, we will consider a Gaussian distributed variable x_G which has the same variance. Due to the same arguments as the ones which preceded Eq. (3.8) it holds that $H_{x_G} \geq H_x$. For the mutual information, we therefore find the following relation

$$MI_{x,s} = H_x - \left\langle H_{x|s}\right\rangle_s \leq H_{x_G} - \left\langle H_{x|s}\right\rangle_s = MI_{ub}\;, \tag{3.39}$$

where MI_{ub} stands for the upper bound for the mutual information. With the Gaussian distribution, $p_G(x) = 1/\sqrt{2\pi\sigma_x^2}\,\exp[-x^2/(2\sigma_x^2)]$, we obtain

$$H_{x_G} = -\int ds\; p(s)\int dx\; p_G(x)\log_2(p_G(x)) = \frac{1}{2}\log_2\left(2\pi\sigma_x^2\right)-\frac{1}{2}\;. \tag{3.40}$$

By inserting Eq. (3.38) and Eq. (3.40) into Eq. (3.39) arrive at the final result of this section, the upper bound for the mutual information,

$$MI_{ub} = \frac{1}{2}\left\langle \log_2\left(\frac{\sigma_x^2}{V(s)}\right)\right\rangle_s\;. \tag{3.41}$$

The expression in Eq. (3.41) constitutes a generalization of the commonly employed upper bound which is discussed in [38] and [62] in the sense, that Eq. (3.41) can account for output variability with an explicit signal dependence. Note however, that we considered here only time-independent variables, while the authors in [38] and [62] discuss time-dependent signals and outputs.

The expression in Eq. (3.41) is an upper bound for the mutual information only if the noise ζ in Eq. (3.36) is Gaussian. In other words, it is an upper bound if for any fixed value of the signal the conditioned variability of the output x is Gaussian distributed. If, additionally, the output x itself is also Gaussian, then MI_{ub} will be equal to the true mutual information. However, if the distribution of the output x differs from the Gaussian distribution, then the upper bound will in general overestimate the true mutual information. The distribution of x is determined by the signal dependence of the functions $M(s)$ and $V(s)$ in Eq. (3.36) and by the distribution of s, which can also be non-Gaussian.

3.3 Numerical measurement of the mutual information for continuously distributed input and output variables

In this section we consider the numerical estimation of the mutual information and discuss some difficulties which arise when continuously distributed variables, e.g. Gaussian distributed variables, are considered. The difficulties which are discussed here are well known in the literature and the reader should refer to other sources for an in-depth discussion of the topic. For a neuroscience-related discussion the reader may consult [61] or [66] and references therein, whereas a discussion of the mathematical intricacies associated with the entropies of continuously distributed variables can be found, for example, in [58] or [90]. However, a reader who is interested in a basic overview of the difficulties which are involved in the numerical measurement of the mutual information or who is interested in a simple algorithm for the estimation of the mutual information between two variables without diving into the complex problem of MI estimation for time-dependent signals may find this chapter useful.

Shannon entropy of a discretized random variable

Consider a system which is subject to a discrete input variable n and generates a discrete output variable N. The entropy of the output is given by

$$H_N = -\sum_i P(N_i) \log_2 P(N_i) \;, \tag{3.42}$$

where $P(N_i)$ is the probability that the variable N attains the state i and where the sum runs over all possible states. The noise entropy of this system, equals the entropy of N for a fixed realization of n averaged over n and reads

$$\Big\langle H_{N|n} \Big\rangle_n = \sum_j P(n_j) \sum_i P(N_i|n_j) \log_2 P(N_i|n_j) \;, \tag{3.43}$$

where $P(n_j)$ is the probability that the variable n attains the state j and where $P(N_i|n_j)$ is the probability, that N is in state i given that n is in state j.

For a continuous variable x with density $p(x)$ there is no clear notion of a state. In order to still be able to compute an entropy, we define a state i as corresponding to the event that a realization of x falls into the interval $[\Delta x(i-\frac{1}{2}), \Delta x(i+\frac{1}{2})]$, where Δx is the discretization which determines the resolution of our measurement of x. If the interval discretization Δx is sufficiently small, then the probability of the state i is given by $P_i = \Delta x \cdot p(i\Delta x)$. If additionally, the variable x is distributed on an infinite interval, we have to choose a lower and upper cutoff (x_{min} and x_{max}), such that the normalization $\sum_i P_i = 1$ is (at least approximately) preserved. The lower and upper cutoff of x then determines the range of i: $x_{min}/\Delta x \leq i \leq x_{max}/\Delta x$. For the entropy we find

$$H_x = -\sum_i \Delta x \; p(i\Delta x) \log_2 \Big(\Delta x \; p(i\Delta x) \Big) \;, \tag{3.44}$$

where again the sum runs over all possible states i. For the noise entropy of x we find

$$\left\langle H_{x|s} \right\rangle_s = -\sum_j \Delta s \; p(j\Delta s) \sum_i \Delta x \; p(i\Delta x|j\Delta s) \log_2 \left(\Delta x \cdot p(i\Delta x|j\Delta s) \right) , \tag{3.45}$$

where $\Delta s \; p(j\Delta s)$ is the probability that a realization of s falls into the interval $[\Delta s(j - \frac{1}{2}), \Delta s(j + \frac{1}{2})]$ and where $\Delta x \; p(i\Delta x|j\Delta s)$ is the probability that x falls into the interval $[\Delta x(i - \frac{1}{2}), \Delta x(i + \frac{1}{2})]$ under the constraint that s fell into the interval $[\Delta s(j - \frac{1}{2}), \Delta s(j + \frac{1}{2})]$.

Because the discretization Δx appear in the argument of the logarithm in Eq. (3.44) and Eq. (3.45), it becomes clear that the entropy of a continuous variable will always depend on the discretization which is used to asses the probabilities P_i. For $\Delta x \to 0$ we find that $H_x \to \infty$ and $\left\langle H_{x|s} \right\rangle_s \to \infty$. Intuitively, this means that for a continuously distributed random variable it is impossible to figure out its exact state with a finite number of binary "yes/no" questions. This fact that the value of the entropy of a continuously distributed variable depends on the discretization will be of importance later, when we discuss the numerical estimation of the mutual information.

For a Gaussian variable, which is measured with a discretization Δx, the entropy is given by [38]

$$H_x = \frac{1}{2} \log_2 \left(\frac{2\pi e \sigma_x^2}{\Delta x^2} \right) . \tag{3.46}$$

Mutual information for discretized random variables

After we convinced ourselves in the previous section that the discretization Δx which is used to define the states x_i of a continuously distributed variable x crucially determines the value of the entropy, let us study how the mutual information between an input variable s and an output variable x depends on the discretization of input and output. For the mutual information we have

$$\begin{aligned} MI =& H_x - \left\langle H_{x|s} \right\rangle_s \\ =& -\sum_i P_i \log_2 (P_i) + \sum_{i,j} P_j P_{i|j} \log_2 \left(P_{i|j} \right) , \end{aligned} \tag{3.47}$$

where in the last term H_x is the full entropy of the output and $\left\langle H_{x|s} \right\rangle_s$ is the noise entropy of x averaged over the signal. As in the previous chapter, state j is attained, whenever a realization of x falls into the interval $[\Delta x(i - \frac{1}{2}), \Delta x(i + \frac{1}{2})]$. Analogously, state j is attained whenever s falls into the interval $[\Delta s(i - \frac{1}{2}), \Delta s(i + \frac{1}{2})]$. For sufficiently small Δs and Δx we find $P_i = \Delta x \; p(i\Delta x)$ and $P_j = \Delta s \; p(j\Delta s)$. Note, that as we convinced ourselves in the previous section, the values of the entropies in the last line of Eq. (3.47) are not independent of Δx and that, in particular, there is no finite limit for $\Delta x \to 0$. Let us now study the dependence of the mutual information on Δx. Therefore,

we consider Eq. (3.47) and use the relation of the probabilities of the states with the probability densities of the continuous variables x and s

$$\begin{aligned} MI &= \sum_{i,j} P_j^s P_{i|j}^x \log_2 \left(\frac{P_{i|j}^x}{P_i^x} \right) \\ &= \sum_{i,j} \Delta s \ \Delta x \ p_s(j\Delta s) p_x(i\Delta x|j\Delta s) \log_2 \left(\frac{p_x(i\Delta x|j\Delta s)}{p_x(i\Delta x)} \right) . \end{aligned} \tag{3.48}$$

We see, that the dependency of the argument of the logarithm on the discretization Δx cancels out. Now, taking in Eq. (3.48) the limit $\Delta x \to dx$ and $\Delta s \to ds$ we can write the sums in the above equations as integrals over x and s

$$MI = \int_{-\infty}^{\infty} ds \int_{-\infty}^{\infty} dx \ p_s(s) p_x(x|s) \log_2 \left(\frac{p_x(x|s)}{p_x(x)} \right) . \tag{3.49}$$

From Eq. (3.48) and Eq. (3.49) we can see that if the input variable s and the output variable x can be characterized by smooth probability distributions (allowing the approximation of the sums by integrals), and if the argument of the logarithm in Eq. (3.49) does not diverge, then the dependence of the mutual information on Δs and Δx vanishes for $\Delta s \to 0$ and $\Delta x \to 0$. Furthermore, the mutual information attains a finite value in contrast to the entropies in the first line of Eq. (3.47). Note, that if the probability densities are not sufficiently smooth or if the argument of the logarithm in Eq. (3.49) diverges, then the mutual information may still depend on the discretization of input and output even for $\Delta s \to 0$ and $\Delta x \to 0$. We will discuss an example of a system that exhibits such a pathological behavior in section 3.4.

Although Eq. (3.49) nicely demonstrates how the dependency of the output entropy and the noise entropy on the discretization cancels out, the probability densities $p_x(x)$, $p_s(s)$ and $p_x(x|s)$ can be computed analytically only for few simple systems. Therefore, in practice the only way to determine the mutual information for a general nonlinear signal processing systems is to discretize the input and output, and to determine the mutual information from the joint probabilities of the discretized input and output. Hereby, as we discussed earlier, it is necessary to carefully choose the discretization of input and output.

In the following we discuss two different routines for the numerical measurement of the mutual information between a continuously distributed input variable s and a continuously distributed output variable x. In order to test the dependence of the numerical estimation of the mutual information on the discretization of the input and output variables we will consider the well-known linear Gaussian channel [38], the output of which is given by

$$x = s + \zeta \ , \tag{3.50}$$

where s is a Gaussian signal variable with variance σ_s^2 and ζ is a Gaussian noise variable

with variance σ_ζ^2. The mutual information between the signal s and the output x in Eq. (3.50) is given by [38]

$$MI = \frac{1}{2}\log_2\left(1+\frac{\sigma_s^2}{\sigma_\zeta^2}\right) . \tag{3.51}$$

The mutual information from Eq. (3.51) is plotted versus the signal standard deviation in Fig. 3.1 for $\sigma_\zeta = 1$ and will provide a reference for the numerical routines which we will introduce in the following.

Numerical method 1: Discretize signal and output and determine the joint probability distribution $P_{i,j}$.

We have already discussed that the entropies of continuously distributed variables x and s crucially depend on the discretization Δs and Δx that are employed to determine the probabilities of the states of the variables and do not have a finite limit for $\Delta x \to 0$ and $\Delta s \to 0$. Because of this divergence, we do not measure the entropies and take the difference as suggested by Eq. (3.47) but rewrite the mutual information in a mathematically equivalent expression

$$\begin{aligned} MI &= -\sum_i P_i \log_2\left(P_i\right) + \sum_{i,j} P_j P_{i|j} \log_2\left(P_{i|j}\right) \\ &= \sum_{i,j} P_{i,j} \log_2\left(\frac{P_{i,j}}{P_i P_j}\right) , \end{aligned} \tag{3.52}$$

where $P_{i,j}$ is the joint probability, that $s \in [\Delta s(j-\frac{1}{2}), \Delta s(j+\frac{1}{2})]$ and at the same time $x \in [\Delta x(i-\frac{1}{2}), \Delta x(i+\frac{1}{2})]$. Note, that we employ the notation P_i in order to indicate probability distributions of discretized continuously distributed variables, while the notation $P(N_i)$, which we employed earlier, indicated the probability distribution of a discrete variable. For the numerical estimation of the probabilities P_i and $P_{i,j}$, the following steps have to be performed:

1. We draw the signal variable s from a predefined distribution.
2. The signal s drives the signal processing system which generates the response x.
3. The output is discretized $x \to x'$ with discretization Δx and the state i of the output is determined from the condition that $x \in [\Delta x(i-\frac{1}{2}), \Delta x(i+\frac{1}{2})]$.
4. We also discretize the input signal $s \to s'$ with a discretization Δs. The state j of the signal is determined from the condition that $s \in [\Delta s(j-\frac{1}{2}), \Delta s(j+\frac{1}{2})]$.

Repeating the above steps N times allows to determine the joint probability $P_{i,j}$ of different pairs of the input and output states $\{i,j\}$. The mutual information is obtained by employing Eq. (3.52), where the marginal probabilities P_i and P_j are obtained by summing the elements of $P_{i,j}$ over j or i respectively.

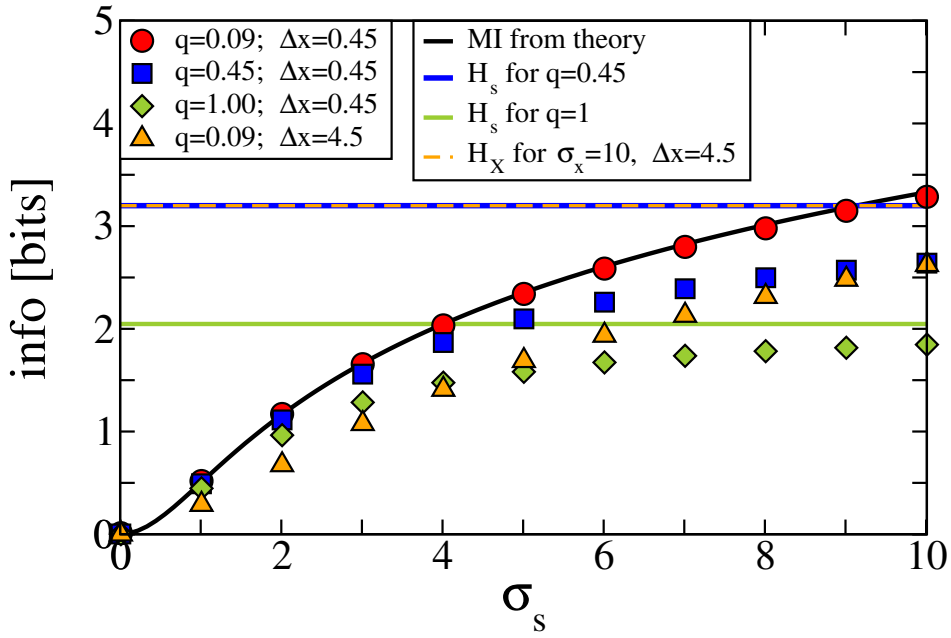

Figure 3.1: Exact mutual information vs σ_s for the linear Gaussian channel measured via numerical method 1. The numerical results for different choices of q and Δx (circles, squares, diamonds and triangles) are compared to the analytical result from Eq. (3.51) (solid black line). The numerical results for the mutual information are always below the entropy of the discretized signal (solid blue line for $q = 0.45$ and solid green line for $q = 1$) and also below the entropy of the discretized output (dashed orange line).

We test the above routine for the linear Gaussian channel, Eq. (3.50), for $\sigma_\zeta = 1$ and for different discretization of input and output. The numerical results for the mutual information are plotted versus the signal standard deviation, σ_s, in Fig. 3.1. The discretization of the Gaussian signal variable is adjusted such that the ratio

$$q = \frac{\Delta s}{\sigma_s} \tag{3.53}$$

remains fixed for all σ_s. For the discretization of the output we choose for every curve a constant binning Δx. The lowest and highest state of the signal is determined numerically by varying the summation limits until the results of the numerical simulations converge and become independent of the choice of the limits. We find that $-3\sigma_s/\Delta s \leq j \leq 3\sigma_s/\Delta s$ is sufficient in all cases which we consider in this thesis. The total number of trials that are used for the measurement of the joint probabilities of input and output is $N = 10^5$. The analytical solution of the mutual information for the linear Gaussian channel is given by Eq. (3.51) and provides a reference for the accuracy of the numerical routine. For the linear Gaussian channel and for the given range of

σ_s the discretization $q = 0.09$ and $\Delta x = 0.45$ (red circles in Fig. 3.1) provides a good match to the analytic theory (black line). Decreasing the discretization of the signal to $q = 0.45$ (blue squares) or to $q = 1$ (green diamonds) systematically underestimates the mutual information. This behavior can be understood by rewriting Eq. (3.47) into

$$\begin{aligned} MI =& H_x - \left\langle H_{x|s} \right\rangle_s \qquad (3.54) \\ =& -\sum_i P_i \log_2 (P_i) + \sum_{i,j} P_j P_{i|j} \log_2 \left(P_{i|j} \right) \\ =& \sum_{i,j} P_{i,j} \log_2 \left(\frac{P_{i,j}}{P_i P_j} \right) , \qquad (3.55) \end{aligned}$$

where from the second last to the last line in the above equation we employed Bayes' theorem, Eq. (1.26). Employing the Bayes' theorem again in Eq. (3.55), the mutual information can be written as

$$\begin{aligned} MI =& -\sum_j P_j \log_2 (P_j) + \sum_{i,j} P_i P_{j|i} \log_2 \left(P_{j|i} \right) \\ =& H_s - \left\langle H_{s|x} \right\rangle_x . \qquad (3.56) \end{aligned}$$

The equivalent expressions for the mutual information in Eq. (3.54) and Eq. (3.56) are simply the consequence of the well known fact, that the mutual information is a symmetric measure with respect to input and output [38]. Considering the fact, that entropies are always positive and in particular $\left\langle H_{s|x} \right\rangle_x > 0$ and $\left\langle H_{x|s} \right\rangle_s > 0$ it follows from Eq. (3.54) and Eq. (3.56) that

$$MI \leq H_x \quad \text{and} \quad MI \leq H_s . \qquad (3.57)$$

In section 3.3 we discussed that the entropies of continuously distributed variables depend on the discretization with which the distributions are measured. From Eq. (3.57) it follows, that by choosing a too large bin size Δs we can systematically underestimate the mutual information. The entropy of a Gaussian signal with discretization $q = \Delta s / \sigma_s$ can be computed from Eq. (3.46). The choices of $q = 0.45$ and $q = 1$ result in $H_s = 3.2$ bits and $H_s = 2.04$ bits, respectively (blue line and green line in Fig. 3.1). Note, that the numerical routine underestimates the mutual information not only for signal standard deviations for which the theory for the mutual information exceeds H_s but even for smaller values. In Fig. 3.1 we find that the numerical routine provides accurate results if q is chosen such that $H_s \approx MI + 1.5$ bits.

Of course, the numerical routine is not only sensitive to q but also to Δx. Increasing the bin size of the output to $\Delta x = 4.5$ (orange triangles in Fig. 3.1) introduces a systematic underestimation of the mutual information and can also be understood by means of Eq. (3.57). For the linear Gaussian channel it is possible to compute H_x in the same way as for the signal by means of Eq. (3.46). For $\Delta x = 4.5$ and $\sigma_x = 10$ this results

in $H_x = 3.2$ bits (orange dashed line in Fig. 3.1). For general nonlinear systems with a non-Gaussian output, one either has to determine H_x numerically or to repeatedly adjust Δx until the numerical result for the mutual information has converged.

Numerical method 2: Determine the conditional probability $P_i(s)$ for a frozen signal, compute the logarithm and average over signals.

An alternative numerical procedure that does not require a discretization of the signal can be deduced by rewriting Eq. (3.52) in another way:

$$\begin{aligned} MI &= \sum_{i,j} P_{i,j} \log_2 \left(\frac{P_{i,j}}{P_i P_j} \right) \\ &= \sum_j P_j \sum_i P_{i|j} \log_2 \left(\frac{P_{i|j}}{P_i} \right) \\ &= \sum_j \Delta s p_s(j\Delta s) \sum_i P_{i|j} \log_2 \left(\frac{P_{i|j}}{P_i} \right) \ . \end{aligned} \tag{3.58}$$

In the limit $\Delta s \to ds$ the first sum can be rewritten as an integration over the signal

$$MI = \int_{-\infty}^{\infty} ds \; p_s(s) \sum_i P_i(s) \log_2 \left(\frac{P_i(s)}{P_i} \right) \ , \tag{3.59}$$

where $P_i(s)$ is now the conditional probability for the output to be in state i for a given signal realization. The multiplication with the probability density $p_s(s)$ and the integration over s in Eq. (3.60) can be considered as an ensemble average over the signal s. Consequently, it is possible to write

$$MI = \left\langle \sum_i P_i(s) \log_2 \left(\frac{P_i(s)}{P_i} \right) \right\rangle_s \ , \tag{3.60}$$

which for a finite ensemble size reads

$$MI = \frac{1}{N_s} \sum_{k=1}^{N_s} \sum_i P_i(s_k) \log_2 \left(\frac{P_i(s_k)}{P_i} \right) \ . \tag{3.61}$$

The variable k in the above equation is now not the state of the signal but a specific realization of the signal variable. The first sum in the last line of Eq. (3.61) runs over all N_s realizations of the signal variable, while the second sum still runs over different states i of the output. The advantage of Eq. (3.61) over Eq. (3.52), which is employed for the numerical routine 1, is that no discretization of the signal is performed and no dependence on Δs is introduced. For the numerical routine 2 the following steps have to be performed:

1. Draw a signal realization s_k.

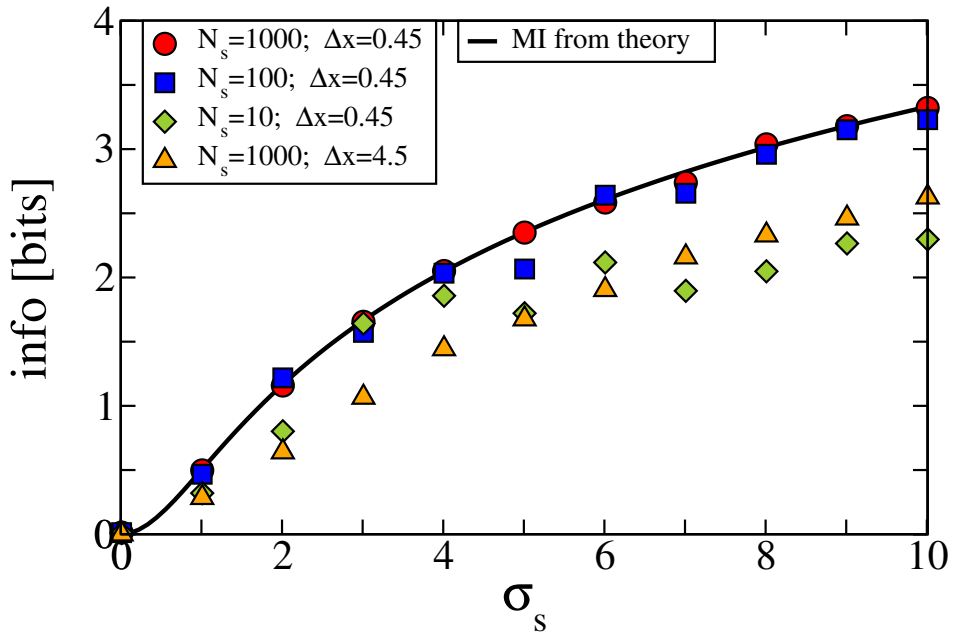

Figure 3.2: Exact mutual information vs σ_s for the linear Gaussian channel measured via numerical method 2. The numerical results for different choices of N_s and Δx (circles, squares, diamonds and triangles) are compared with the analytical result from Eq. (3.51) (solid black line).

2. Determine the response x to the input s_k
3. Discretize $x \to x'$ and determine the state i according to the condition that $x \in [\Delta x(i - \frac{1}{2}), \Delta x(i + \frac{1}{2})]$
4. repeat steps 2 and 3 and determine the conditional probability $P_i(s_k)$ for the frozen signal realization s_k.

Repeating the above steps N_s times allows to determine the two dimensional array $P_i(s_k)$. The marginal probability distribution P_i is obtained from the ensemble average $P_i = \langle P_i(s_k) \rangle_{s_k}$. Finally, the mutual information is computed from Eq. (3.61). The advantage of this numerical method is that it is not necessary to choose a somehow arbitrary discretization of the signal. After all, the signal processing system is subject to the continuously distributed signal and not to a discretized version of it. However, the avoidance of the discretization comes at a cost, as we have added an additional parameter, the ensemble size N_s. In order to test the numerical routine 2 for different parameter sets we apply it to the linear Gaussian channel, Eq. (3.50), with $\sigma_\zeta = 1$. The numerical results for the mutual information are plotted versus σ_s in Fig. 3.2. As before, the highest and lowest state of the output is determined numerically by varying the summation limits until the results of the numerical simulations converge and become independent of the choice of the limits. The total number of trials that are used to

determine the conditional probabilities $P_i(s_k)$ is $N = 1000$. Note, that the conditional probability has to be determined for every realization of the signal, resulting in a total of $N \cdot N_s$ simulations of the signal processing system. For $N_s = 1000$ (red dots in Fig. 3.2) the match with the analytic theory, Eq. (3.51), is quite good. Reducing the ensemble size to $N_s = 100$ (blue squares in Fig. 3.2) or $N_s = 10$ (green diamonds in Fig. 3.2) results not only in a larger statistical variation around the theoretical curve but also to a systematic underestimation of the mutual information. Unfortunately, no simple criterion can be found which would predict the minimal ensemble size that is required to attain a measurement with a certain precision. In general, the optimal ensemble size will be different for different signal processing systems depending on the amount and nature of the intrinsic noise and the type of the signal processing system's nonlinearity. Finally, an increase of the discretization of the output to $\Delta x = 4.5$ (orange triangles in Fig. 3.2) results in an underestimation of the mutual information due to Eq. (3.57) and is the same effect that was already observed for the numerical routine 1. Note, that the difference between the numerical routine 1 and 2 is not only the analysis of the data (measurement of joint probabilities vs conditional probabilities) but also the generation of the data. While for the numerical routine 1 a different signal realization is drawn for every simulation of the signal processing system (N simulations in total), the numerical routine 2 requires N simulations of the signal processing system for every one of the N_s realizations of the signal ($N \cdot N_s$ simulations in total).

Finally, we compare the numerical results from routine 1 with $q = 0.45$ and $N = 100000$ (red circles in Fig. 3.3) with the results from routine 2 with $N_s = 1000$ and $N = 1000$ (blue squares in Fig. 3.3). For $\sigma_s \leq 6$ the results from both routines show a good agreement with the theoretical prediction for the mutual information for the linear Gaussian channel, Eq. (3.51) (black line in Fig. 3.3). For $\sigma_s > 6$, however, we find that the results from the numerical routine 1 show a slightly better agreement with the theoretical prediction than the results from the numerical routine 2. Furthermore, routine 1 requires only $N = 10^5$ simulations of the signal processing system, which is an order of magnitude smaller than the $N_s \cdot N = 10^6$ simulations that are required for the numerical routine 2. Consequently, for the remainder of this chapter we will employ routine 1 whenever a numerical measurement of the mutual information is required.

3.4 Nonlinear decoding of a signal variable from the output of a Gaussian model with signal-dependent mean and variance

In order to gain some intuition about how the mutual information and its upper and lower bounds behave for a simple nonlinear system, we consider a Gaussian model, the output of which is given by

$$x = M(s) + \sqrt{V(s)}\,\zeta\ . \tag{3.62}$$

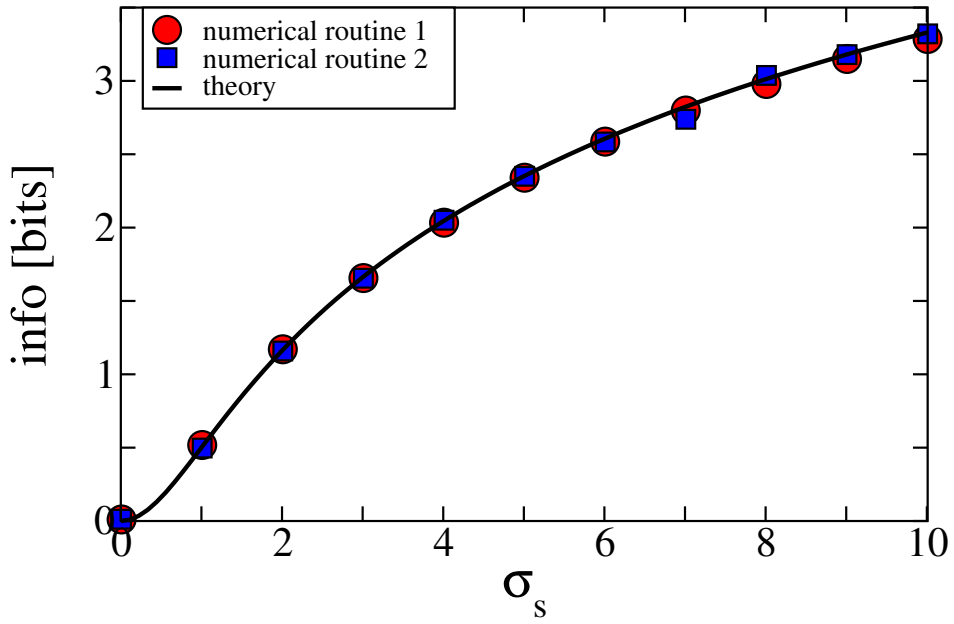

Figure 3.3: Comparison of the numerical results for the mutual information of a linear Gaussian channel from the numerical method 1 and the numerical method 2. The analytical curve for the mutual information is given by Eq. (3.51) (solid black line). The numerical data from method 1 (Fig. 3.1) that fits the analytical curve best is displayed by the red circles ($q = 0.09$, $\Delta x = 0.45$, $N = 10^5$). The numerical data from method 2 (Fig. 3.2) that fits the analytical curve best is displayed by the blue squares ($N_s = 10^3$, $\Delta x = 0.45$, $N = 10^3$).

The signal s in the above equation is a Gaussian distributed input signal with variance σ_s^2 and ζ is a Gaussian distributed noise with variance

$$\sigma_\zeta^2 = 1 \; . \tag{3.63}$$

For the remainder of the chapter we will assume that

$$\langle s \rangle = 0 \; , \quad \langle \zeta \rangle = 0 \; . \tag{3.64}$$

The function $M(s)$ in Eq. (3.62) is identical to the mean output for a fixed signal realization,

$$\langle x \rangle_\zeta = M(s) + \sqrt{V(s)}\, \langle \zeta \rangle_\zeta = M(s) \; . \tag{3.65}$$

In the above equation, the angular brackets, $\langle . \rangle_\zeta$, with the subscript ζ indicate an average of the variables inside the bracket, where only the noise ζ is varied from trial to trial but the signal is fixed at a deterministic value. In contrast, angular brackets without a subscript, $\langle . \rangle$, will be used to indicate simultaneous trial averaging over the signal s and

the noise ζ.

The function $V(s)$ in Eq. (3.62) constitutes the variance of the output for a fixed signal realization

$$\left\langle \Delta x^2 \right\rangle_\zeta = \left\langle (x - \langle x \rangle_\zeta)^2 \right\rangle_\zeta = V(s) \left\langle \zeta\zeta \right\rangle_\zeta = V(s) \; . \tag{3.66}$$

In an experimental setting in which the function $M(s)$ is not known beforehand, Eq. (3.65) indicates how it is possible to measure $M(s)$. The measurement is performed by averaging the output of a system for a fixed value, s, of the input signal (frozen signal) over different trials in which the intrinsic noise is independent from trial to trial. Analogously, it is also possible to determine $V(s)$, but instead of simply averaging the output for a fixed signal s, one has to compute the variance of the output for fixed realizations of the signal. Repeating the measurement of $M(s)$ and $V(s)$ for different values of the signal, allows to determine the full functions $M(s)$ and $V(s)$. Because of their correspondence to trial averages for frozen signals, in the following, we will refer to $M(s)$ as the signal-dependent mean of the output and $V(s)$ as the signal-dependent variance of the output.

For the output in Eq. (3.62), the squared linear and nonlinear correlation coefficients Eq. (3.20), Eq. (3.28) and Eq. (3.29) become

$$\rho^2_{x,s} = \frac{\langle M(s) \cdot s \rangle^2}{\sigma_s^2 \sigma_x^2} \; , \tag{3.67}$$

$$\rho^2_{x^2,s} = \frac{\left(\langle M^2(s) \cdot s \rangle + \langle V(s) \cdot s \rangle \right)^2}{\sigma_s^2 \sigma_{x^2}^2} \; , \tag{3.68}$$

$$\rho^2_{x^2,x} = \frac{\left(\langle M^3(s) \rangle - \langle M(s) \rangle \langle M^2(s) \rangle + 3 \langle M(s) V(s) \rangle - \langle M(s) \rangle \langle V(s) \rangle \right)^2}{\sigma_x^2 \sigma_{x^2}^2} \; , \tag{3.69}$$

where

$$\sigma_x^2 = \Big\langle M^2(s) \Big\rangle - \Big\langle M(s) \Big\rangle^2 + \Big\langle V(s) \Big\rangle \; ,$$

and

$$\sigma_{x^2}^2 = \Big\langle M^4(s) \Big\rangle - \Big\langle M^2(s) \Big\rangle^2 + 6 \Big\langle M^2(s) V(s) \Big\rangle - \Big\langle M^2(s) \Big\rangle \Big\langle V(s) \Big\rangle + 3 \Big\langle V^2(s) \Big\rangle - \Big\langle V(s) \Big\rangle^2 .$$

The exact mutual information between the signal s and the Gaussian output variable x is given by

$$MI = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} ds \, dx \; p(s) p(x|s) \log_2 \left(\frac{p(x|s)}{\int_{-\infty}^{\infty} ds' p(s') p(x|s')} \right) \; , \tag{3.70}$$

where

$$p(s) = \frac{1}{\sqrt{2\pi\sigma_s^2}} \exp\left[-\frac{s^2}{2\sigma_s^2}\right] \tag{3.71}$$

and

$$p(x|s) = \frac{1}{\sqrt{2\pi V(s)}} \exp\left[-\frac{(x - M(s))^2}{2V(s)}\right] . \tag{3.72}$$

For general nonlinear $M(s)$ and $V(s)$ it is not possible to find a simple analytical expression for the exact mutual information. Therefore, for the remainder of this chapter, we will compute the exact mutual information for the Gaussian model via a numerical evaluation of the integrals in Eq. (3.70).

For the system in Eq. (3.62), we now proceed with the study of the exact mutual information and its upper and lower bounds for different choices of $M(s)$ and $V(s)$.

Linear Gaussian channel

One simple choice for the signal-dependent mean and variance in Eq. (3.62) is

$$M(s) = s \quad \text{and} \quad V(s) = 1 , \tag{3.73}$$

which leads to the linear Gaussian channel, Eq. (3.50). The functions $M(s)$ and $V(s)$ from Eq. (3.73) are plotted in Fig. 3.4a and Fig. 3.4b respectively. In Fig. 3.4c we plot the distribution of the input signal normalized by the maximum of the respective distribution for three different values of the standard deviation. Figure 3.4 illustrates the simple fact, that the functions $M(s)$ and $V(s)$ are linear. For the linear Gaussian channel, the standard deviation of the signal does not affect the shape of the distribution of x, which is always Gaussian. The mutual information of the linear Gaussian channel was already computed in Eq. (3.51) and can be rewritten as

$$MI = \frac{1}{2}\log_2\left(1 + \frac{\sigma_s^2}{\sigma_\zeta^2}\right) = -\frac{1}{2}\log_2\left(1 - \frac{\sigma_s^2}{\sigma_x^2}\right) , \tag{3.74}$$

where we employed that for the linear Gaussian channel $\sigma_x^2 = \sigma_s^2 + \sigma_\zeta^2$ and that $\sigma_\zeta^2 = 1$. The squared linear correlation coefficient, Eq. (3.67), between the signal, s, and the output from Eq. (3.50) can be easily computed and yields

$$\rho_{x,s}^2 = \frac{\sigma_s^2}{\sigma_x^2} . \tag{3.75}$$

Inserting Eq. (3.75) into the equation for the linear lower bound, Eq. (3.21), results in

$$MI_{lb}^{lin} = -\frac{1}{2}\log_2\left(1-\frac{\sigma_s^2}{\sigma_x^2}\right) , \tag{3.76}$$

which is the same result as for the exact mutual information, Eq. (3.74). This result was of course to be expected because for a linear system with additive intrinsic noise the optimal linear reconstruction already extracts all the information that is present in the output.

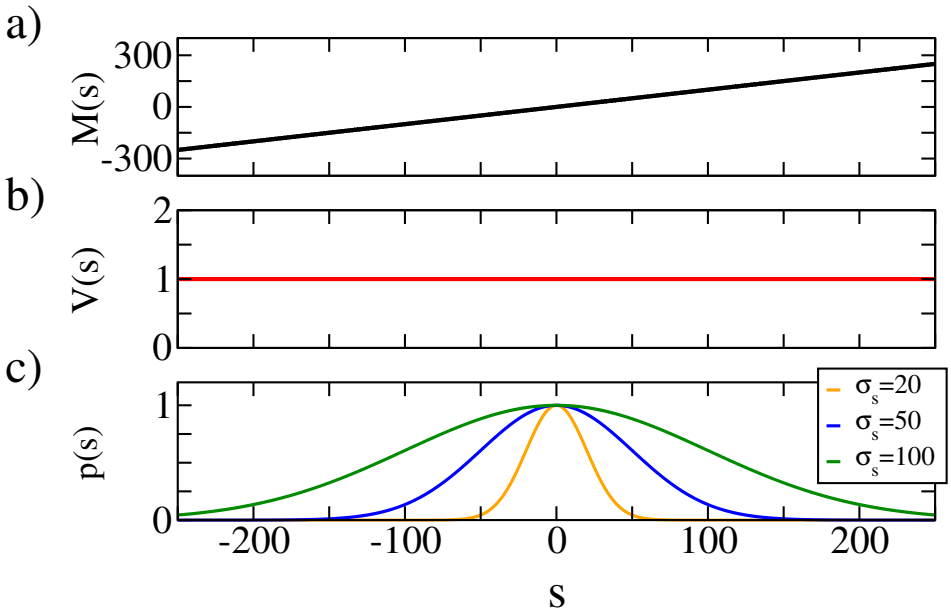

Figure 3.4: Mean, variance, and signal distributions for the linear Gaussian channel. a) $M(s)$, as a function of the signal-value. $M(s)$ corresponds to the output of the linear Gaussian channel for a fixed signal value averaged over independent realizations of the noise. b) $V(s)$ as a function of the signal-value. $V(s)$ corresponds to the variance of the output of the linear Gaussian channel for a fixed signal value averaged over independent realizations of the noise. c) Signal distributions, $p(s)$, as functions of the signal-value for different σ_s rescaled such that $p(s=0)=1$. The signal distributions are indicators of the frequency with which the stochastic signal variable s attains certain values.

For the computation of the quadratic lower bound we first consider the cross correlation between the signal s and the squared output x^2

$$\Big\langle \Big(x^2-\langle x^2\rangle\Big)\Big(s-\langle s\rangle\Big)\Big\rangle = \Big\langle \Big(s^2+\zeta^2-2s\zeta-\sigma_s^2-1\Big)\Big(s-\langle s\rangle\Big)\Big\rangle \ . \tag{3.77}$$

Employing in Eq. (3.77) that the signal follows a Gaussian distribution with zero mean

and, hence, all odd moments of s equal to zero, leads to

$$\Big\langle \Big(x^2 - \langle x^2 \rangle\Big)\Big(s - \langle s \rangle\Big)\Big\rangle = 0 \tag{3.78}$$

and, consequently, implies for the squared nonlinear correlation coefficient

$$\rho^2_{x^2,s} = \frac{\big\langle (x^2 - \langle x^2 \rangle)(s - \langle s \rangle)\big\rangle^2}{\sigma^2_{x^2}\sigma^2_s} = 0 \; . \tag{3.79}$$

Inserting Eq. (3.79) into the equation for the quadratic lower bound, Eq. (3.33) leads to

$$MI^{quad}_{lb} = -\frac{1}{2}\log_2\left(1 - \frac{\sigma^2_s}{\sigma^2_x}\right) \; , \tag{3.80}$$

which is again the same result as for the exact mutual information given by Eq. (3.74). This result was also to be expected, because if all the information that is contained in the output is already recovered by the optimal linear reconstruction, then the quadratic reconstruction cannot provide any additional information.

Finally, the upper bound for the mutual information, Eq. (3.41), results for the linear Gaussian channel in

$$\begin{aligned} MI_{ub} &= \frac{1}{2}\log_2\left(\frac{\sigma^2_x}{1}\right) = -\frac{1}{2}\log_2\left(\frac{\sigma^2_x - \sigma^2_s}{\sigma^2_x}\right) \\ &= -\frac{1}{2}\log_2\left(1 - \frac{\sigma^2_s}{\sigma^2_x}\right) \; . \end{aligned} \tag{3.81}$$

As expected, due to the fact that the output of the linear Gaussian channel is Gaussian distributed, Eq. (3.81) in this case equals the exact mutual information that is given by Eq. (3.74). The different measures of information transmission for the linear Gaussian channel are plotted versus the signal standard deviation in Fig. 3.5. As discussed above, for the special case of the linear Gaussian channel the exact mutual information and its lower and upper bounds coincide. In particular, the bounds and the mutual information grow for increasing signal strength without any saturation.

Gaussian model with a nonlinear signal dependence of the mean

We now consider a Gaussian model with a constant variance and a weakly nonlinear signal dependence of the mean:

$$M(s) = s + \alpha s^2 \quad \text{and} \quad V(s) = 1 \; , \tag{3.82}$$

where α is a constant coefficient. For the output variable x this leads to

$$x = s + \alpha s^2 + \zeta \; . \tag{3.83}$$

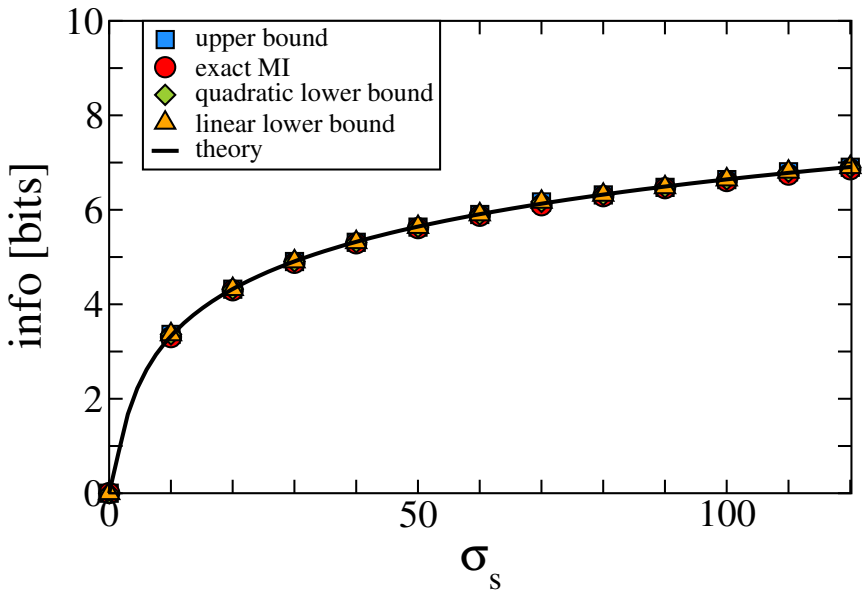

Figure 3.5: Mutual information (red circles), linear lower bound (orange triangles), quadratic lower bound (green diamonds) and the upper bound (blue squares) for the linear Gaussian channel vs the signal standard deviation. The theory (black line) is computed via Eq. (3.74).

The functions $M(s)$ and $V(s)$ are plotted as functions of s in Fig. 3.6a,b for $\alpha = 0.001$. The normalized signal distributions for $\sigma_s = 20,\ 50,\ 100$ are plotted in Fig. 3.6c. For the displayed range of signal values, the function $M(s)$ exhibits only a weakly nonlinear behavior and by the naked eye it is only hardly distinguishable from a linear function. Of course, for larger σ_s the signal would sample a wider ranger of values, and the nonlinear behavior of $M(s)$ would be more pronounced. However, for $\sigma_s \leq 100$ the signal samples the functions $M(s)$ and $V(s)$ mostly within the range for which $M(s)$ appears almost linear (cf. with $M(s)$ in Fig. 3.4).

Let us now consider the lower bounds for the mutual information. By using Eq. (3.83) and employing that $\langle s \rangle = \langle \zeta \rangle = 0$ and $\sigma_\zeta^2 = 1$ the squared linear correlation coefficient, Eq. (3.20), yields

$$\rho_{x,s}^2 = \frac{\sigma_s^2}{2\alpha^2\sigma_s^4 + \sigma_s^2 + 1} \,. \tag{3.84}$$

For the squared correlation coefficient between the signal and the squared output,

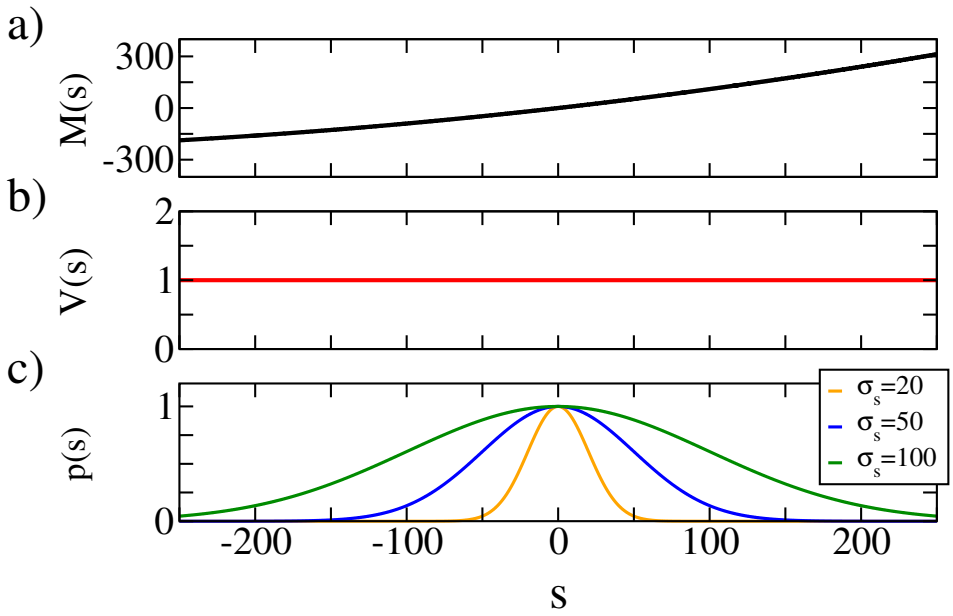

Figure 3.6: Mean, variance, and signal distributions for the Gaussian model with a nonlinear signal dependence of the mean for $\alpha = 0.001$. a) Function $M(s)$ vs signal-value; b) Function $V(s)$ vs signal-value; c) Signal distributions, $p(s)$, as functions of the signal-value for different σ_s rescaled such that $p(s = 0) = 1$.

Eq. (3.28), one finds

$$\rho^2_{x^2,s} = \frac{18\alpha^2\sigma_s^6}{48\alpha^4\sigma_s^8 + 42\alpha^2\sigma_s^6 + \sigma_s^4(1 + 6\alpha^2) + 2\sigma_s^2 + 1} , \tag{3.85}$$

and for the squared correlation coefficient between the output and the squared output, Eq. (3.29), the result reads

$$\rho^2_{x^2,x} = \frac{2\alpha^2\sigma_s^4(1 + 4\sigma_s^2 + 6\alpha^2\sigma_s^4)^2}{(1 + \sigma_s^2 + 2\alpha^2\sigma_s^4)(1 + \sigma_s^4 + 42\alpha^2\sigma_s^6 + 48\alpha^4\sigma_s^8 + 2(\sigma_s^2 + 3\alpha^2\sigma_s^4))} . \tag{3.86}$$

For the upper bound for the mutual information, Eq. (3.41), one finds

$$MI_{ub} = \frac{1}{2}\log_2\left(2\alpha^2\sigma_s^4 + \sigma_s^2 + 1\right) . \tag{3.87}$$

The linear lower bound is computed by inserting Eq. (3.84) into Eq. (3.21), while the quadratic lower bound is computed by inserting Eqs. (3.84-3.86) into Eq. (3.33). The exact mutual information, is computed via numerically evaluating Eq. (3.70) and is plotted versus σ_s in Fig. 3.7 together with its upper and lower bounds for $\alpha = 0.001$. Note, that for the signal range indicated by the signal distributions in Fig. 3.6c, the signal-dependent mean in Fig. 3.6a shows only a weakly nonlinear behavior which by the

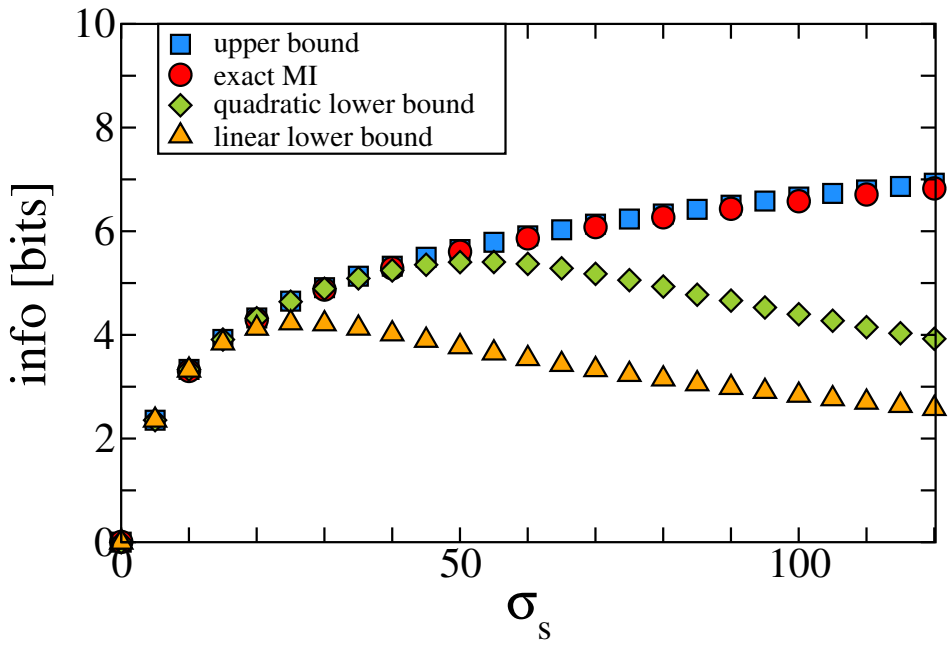

Figure 3.7: Mutual information (red circles), linear lower bound (orange triangles), quadratic lower bound (green diamonds) and the upper bound (blue squares) for the Gaussian model with a nonlinear signal dependence of the mean vs the signal standard deviation for $\alpha = 0.001$.

naked eye is only hardly distinguishable from the linear case. Nevertheless, the behavior of the lower bounds for increasing σ_s is markedly different from what we have observed for the linear Gaussian channel (cf. Fig. 3.5). In Fig. 3.7, the linear lower bound (orange triangles) reaches a maximum at around $\sigma_s = 20$ and decreases for larger signal strengths. The quadratic lower bound (green diamonds) reaches a maximum at around $\sigma_s = 50$ and decreases for larger signal strength similarly to the linear lower bound. The exact mutual information (red circles) increases monotonically for increasing signal strength in a simmilar manner as for the linear Gaussian channel and is closely matched by the upper bound (blue squares). For $\sigma_s \leq 20$ both lower bounds match the exact mutual information closely and increase monotonically for increasing signal strength, comparable to the behavior of the linear Gaussian channel. This behavior can be explained by the small value of α and by the dominance of the linear term in s in Eq. (3.83) for small standard deviations of the signal. For $20 \leq \sigma_s \leq 50$, in which case the linear lower bound already shows strong deviations from the exact mutual information, the quadratic lower bound still captures almost all of the information. From a reconstruction perspective, the linear lower bound indicates that for $\sigma_s = 60$ the optimal linear filter extracts around 3.5 bits of information. For the same signal standard deviation the optimal quadratic reconstruction increases the extracted information by around 53% to around 5.4 bits. Although for $\sigma_s \geq 50$ the distance between both lower bounds and the exact mutual

information increases steadily for increasing signal standard deviation, the quadratic lower bound lies always above the linear lower bound.

In order to gain an intuition about the behavior of the lower bounds for strong signals, we consider the limit $\sigma_s \to \infty$ for the squared correlation coefficients, Eqs. (3.84-3.86), which results in

$$\lim_{\sigma_s \to \infty} \rho^2_{x,s} = 0 \,, \tag{3.88}$$

$$\lim_{\sigma_s \to \infty} \rho^2_{x^2,s} = 0 \,, \tag{3.89}$$

$$\lim_{\sigma_s \to \infty} \rho^2_{x^2,x} = \frac{3}{4} \,. \tag{3.90}$$

For the linear lower bound, Eq. (3.21), and the quadratic lower bound, Eq. (3.33), the above three equations lead to

$$\lim_{\sigma_s \to \infty} MI^{lin}_{lb} = 0 \,, \tag{3.91}$$

$$\lim_{\sigma_s \to \infty} MI^{quad}_{lb} = 0 \,. \tag{3.92}$$

In contrast to the upper bound, Eq. (3.87), which grows without bounds for increasing signal variance, the linear and the quadratic lower bounds decay to zero in the limit of infinitely strong signals. While we do not know the exact behavior of the true mutual information with respect to σ_s, the behavior of the upper bound hints at the possibility that the mutual information may also increase without bounds for increasing σ_s. This implies that the lower bounds can only provide a reasonable estimate for the true mutual information for small and intermediate signal standard deviations.

In conclusion, we find for the Gaussian model with a weakly nonlinear signal dependence: i) the quadratic lower bound provides a better estimate for the exact mutual information than the linear lower bound for intermediate σ_s; ii) the optimal quadratic filter recovers significantly more information from the output than the optimal linear filter for intermediate σ_s.

Gaussian model with pure noise coding

Now we consider a Gaussian model for which the mean is signal independent and only the variance exhibits a signal dependence:

$$M(s) = 0 \,, \quad V(s) = (1 + \beta s)^2 \,, \tag{3.93}$$

where β is a constant parameter. For the output x, Eq. (3.93) leads to

$$x = (1 + \beta s)\, \zeta \,. \tag{3.94}$$

In the following we will refer to the model in the above equation as the Gaussian model with pure noise coding. Note, that our setup is different from what is referred to as

variance coding (e.g. in [91]), where the intrinsic noise of a model neuron is modulated by the signal. Such a variance coding may lead to a signal-dependent variability of the output but will in general also lead to a signal dependence of the mean output. Here, the term noise coding refers solely to the signal dependence of the output variability, which is given by $V(s)$.

The functions $M(s)$ and $V(s)$ are plotted in Fig. 3.8a and Fig. 3.8b, respectively, for $\beta = 0.01$. The signal distributions that indicate which values are most often assumed by the signal are plotted in Fig. 3.8c for different σ_s. For $\sigma_s = 20$, approximately $99,7\%$ of the signal values lie in the interval $[-60, 60]$. Consequently, the function $V(s)$ is predominantly sampled within an interval within which the function is invertible. This means that a fixed signal value (a frozen signal) can be uniquely determined from a measurement of the signal-dependent variance $V(s)$. For $\sigma_s = 100$, the values which are assumed by the stochastic signal are distributed over a much larger interval and the function $V(s)$ is sampled over an interval on which it is not invertible anymore. In this case, a measurement of $V(s)$ does not allow to uniquely determine the value of a frozen input signal anymore.

For the linear correlation coefficient we find

$$\rho^2_{x,s} = 0 \,, \tag{3.95}$$

which leads to

$$MI^{lin}_{lb} = 0 \tag{3.96}$$

for all σ_s. Hence, the linear reconstruction extracts no information from a source with pure noise coding regardless of the signal strength.

For the nonlinear coefficients we find

$$\rho^2_{x^2,s} = \frac{2\beta^2\sigma_s^2}{1 + 8\beta^2\sigma_s^2 + 4\beta^4\sigma_s^4} \tag{3.97}$$

and

$$\rho^2_{x^2,x} = 0 \,. \tag{3.98}$$

Consequently, for the quadratic lower bound we obtain

$$MI^{quad}_{lb} = -\frac{1}{2}\log_2\left(1 - \frac{2\beta^2\sigma_s^2}{1 + 8\beta^2\sigma_s^2 + 4\beta^4\sigma_s^4}\right) \,, \tag{3.99}$$

which attains a maximum at $\sigma_s = 1/(\sqrt{2}\cdot|\beta|)$. Inserting the optimal σ_s into Eq. (3.99) leads to

$$MI^{quad}_{lb,max} = -0.5\,(\log_2(5) - \log_2(6)) \approx 0.13 \,. \tag{3.100}$$

For the optimal signal strength, the quadratic reconstruction recovers around 0.13 bits

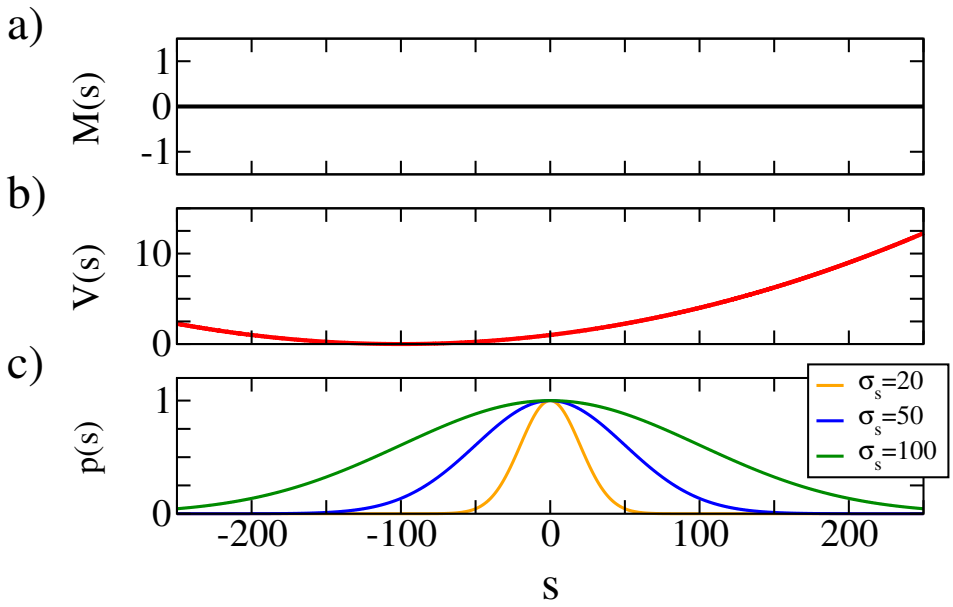

Figure 3.8: Mean, variance, and signal distributions for the Gaussian model with pure noise coding for $\beta = 0.01$. a) Function $M(s)$ vs signal-value; b) Function $V(s)$ vs signal-value; c) Signal distributions, $p(s)$, as functions of the signal-value for different σ_s rescaled such that $p(s = 0) = 1$. The signal distributions indicate the range of values which are predominantly assumed by the stochastic signal. This in turn, determines which parts of $M(s)$ and $V(s)$ are predominantly sampled by the signal.

of information from the output and is completely independent of the parameter β. In the limit of a strong signal the nonlinear coefficient from Eq. (3.97) reduces to

$$\lim_{\sigma_s \to \infty} \rho_{x^2,s} = 0 \,, \tag{3.101}$$

which leads to

$$\lim_{\sigma_s \to \infty} MI_{lb}^{quad} = 0 \,. \tag{3.102}$$

Interestingly, while the linear lower bound, Eq. (3.96), is zero regardless of the signal strength, the quadratic lower bound extracts up to 0.13 bits of information for a signal of intermediate strength and decays to zero for large σ_s. The linear lower bound, Eq. (3.96), and the quadratic lower bound, Eq. (3.99) and the exact mutual information, Eq. (3.70), are plotted in Fig. 3.9 for $\beta = 0.01$. The linear lower bound (orange triangles) is zero for all σ_s as predicted by Eq. (3.96). The quadratic lower bound (green diamonds) increases for increasing σ_s until it reaches a peak at $\sigma_s = 70$ and then it again slowly decays for increasing σ_s. For $\sigma_s \leq 20$ the quadratic lower bound closely follows the exact mutual information (red circles). For $\sigma_s > 20$ the quadratic lower bound starts to significantly deviate from the exact mutual information, but note that even for $\sigma_s = 150$ the quadratic

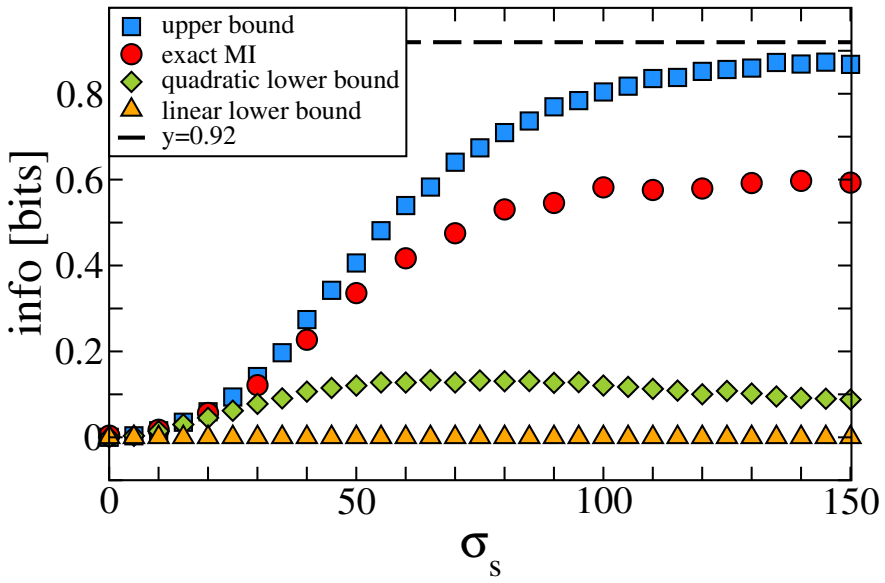

Figure 3.9: Mutual information (red circles), linear lower bound (orange triangles), quadratic lower bound (green diamonds) and the upper bound (blue squares) for the Gaussian model with pure noise coding for $\beta = 0.01$. The limit of the upper bound for $\sigma_s \to \infty$ (dashed black line) is given by Eq. (3.109).

reconstruction still recovers around 0.1 bits of information from the output in marked contrast to the 0 bits of information that are recovered by the linear reconstruction.

In order to be able to put the 0.1 bits of information into perspective, let us consider a linear Gaussian channel, for which the exact mutual information can be related to the linear correlation coefficient between input and output via Eq. (3.74) and Eq. (3.75). Inserting Eq. (3.75) into Eq. (3.74) and solving for $\rho_{x,s}$ we arrive for the linear Gaussian channel at

$$\rho_{x,s} = \sqrt{1 - e^{-2 \cdot MI \cdot \ln(2)}} \; . \tag{3.103}$$

Inserting $MI = 0.1$ into the above equation results in $\rho_{x,s} \approx 0.36$. Consequently, a linear Gaussian channel that transmits 0.1 bits of information exhibits a rather substantial linear correlation between input and output of about 0.36. We therefore argue that a quadratic reconstruction that extracts up to 0.13 bits of information is a major improvement in comparison to the linear reconstruction, which cannot extract any information from a channel with pure noise coding.

Now we consider the upper bound, Eq. (3.41), which for the Gaussian model with pure

noise coding is given by

$$MI_{ub} = \frac{1}{2}\left\langle \log_2\left(\frac{1+\beta^2\sigma_s^2}{(1+\beta s)^2}\right)\right\rangle_s \ . \tag{3.104}$$

Writing the ensemble average in Eq. (3.104) as an integration, we find

$$\begin{aligned} MI_{ub} &= -\frac{1}{2}\int_{-\infty}^{\infty} ds\ p_s(s)\log_2\left(\frac{(1+\beta s)^2}{1+\beta^2\sigma_s^2}\right) \\ &= -\frac{1}{2}\frac{1}{\sqrt{2\pi\sigma_s^2}}\int_{-\infty}^{\infty} ds\ e^{-\frac{s^2}{2\sigma_s^2}}\log_2\left(\frac{(1+\beta s)^2}{1+\beta^2\sigma_s^2}\right) \ . \end{aligned} \tag{3.105}$$

After a variable transformation, the above equation transforms into

$$MI_{ub} = -\frac{1}{2}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} ds\ e^{-\frac{s^2}{2}}\log_2\left(\frac{\frac{1}{\sigma_s^2}+\frac{2\beta s}{\sigma_s}+\beta^2 s^2}{\frac{1}{\sigma_s^2}+\beta^2}\right) \ . \tag{3.106}$$

In the limit $\sigma_s \to \infty$ Eq. (3.106) reduces to

$$\lim_{\sigma_s\to\infty} MI_{ub} = -\frac{1}{2}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} ds\ e^{-\frac{s^2}{2}}\log_2\left(s^2\right) \ . \tag{3.107}$$

Using that [92, p. 197]

$$\int_0^{\infty} dx\ e^{-x^2}\ln(x) = -\frac{\sqrt{\pi}}{4}\left(\gamma + 2\ln(2)\right) \ , \tag{3.108}$$

where $\gamma \approx 0.58$ is the Euler-Mascheroni constant [93], we can simplify Eq. (3.107) into

$$\lim_{\sigma_s\to\infty} MI_{ub} = \frac{1}{2}\left(1+\frac{\gamma}{\ln(2)}\right) \approx 0.92 \ . \tag{3.109}$$

The upper bound is represented by the blue squares in Fig. 3.9. In stark contrast to the lower bounds, which decay to zero for increasing signal variance, the upper bound converges to a finite value of around 0.92 bits (dashed line in Fig. 3.9). This implies that for the Gaussian model with noise coding, also the mutual information (red circles in Fig. 3.9) does not grow without bounds but has to be lower than 0.92 bits even in the limit of a signal with infinite variance. This behavior stands in marked contrast to the linear Gaussian channel and the Gaussian model with a nonlinear signal dependence of the mean, for which the mutual information seems to grow without bounds for increasing signal variance.

In conclusion, we find that for the Gaussian model with pure noise coding: i) The

quadratic lower bound provides a reasonable estimate for the exact mutual information if the function $V(s)$ is sampled by the signal predominantly in a region, where $V(s)$ is invertible. For $\beta = 0.01$ this corresponds to $\sigma_s < 20$. ii) The optimal quadratic filter is able to recover information from a channel with pure noise coding in contrast to the linear reconstruction, which does not recover any information for all σ_s. iii) The Gaussian model with pure noise coding constitutes a nice example of a system for which the mutual information does not increase with increasing standard deviation of the Gaussian signal.

Piecewise linear Gaussian model

So far we only studied Gaussian models with smooth nonlinearities. Now we will consider a piecewise linear Gaussian model with

$$M(s) = s \cdot \Theta(s) \;, \quad V(s) = \Theta(s) \;. \tag{3.110}$$

For the output x, Eq. (3.62), the above choice of $M(s)$ and $V(s)$ leads to

$$x = (s + \zeta)\, \Theta(s) \;, \tag{3.111}$$

where Θ is the Heaviside step function. The above choice of $M(s)$ and $V(s)$ is motivated by the behavior of neural models, for which the count cannot attain negative values. Consequently, strong negative signals cannot be transmitted, because they essentially silence the neuron. This behavior is mimicked by the Gaussian model in Eq. (3.111). The functions $M(s)$ and $V(s)$ are plotted in Fig. 3.10a and Fig. 3.10b, respectively.

The squared linear correlation coefficient, Eq. (3.67), for the piecewise linear Gaussian model results in

$$\rho_{x,s}^2 = \frac{1}{2\left(1 - \frac{1}{\pi} + 2\frac{1}{\sigma_s^2}\right)} \;. \tag{3.112}$$

which in the limit $\sigma_s \to \infty$ reduces to

$$\lim_{\sigma_s \to \infty} \rho_{x,s}^2 = \frac{1}{2\left(1 - \frac{1}{\pi}\right)} \approx 0.733 \;. \tag{3.113}$$

After inserting the limit from the above equation into Eq. (3.21) we obtain a finite value for the linear lower bound in the limit of strong signals,

$$\lim_{\sigma_s \to \infty} MI_{lb}^{lin} \approx 0.95 \text{ bits} \;. \tag{3.114}$$

For the squared correlation coefficient between the input and the squared output, Eq. (3.68),

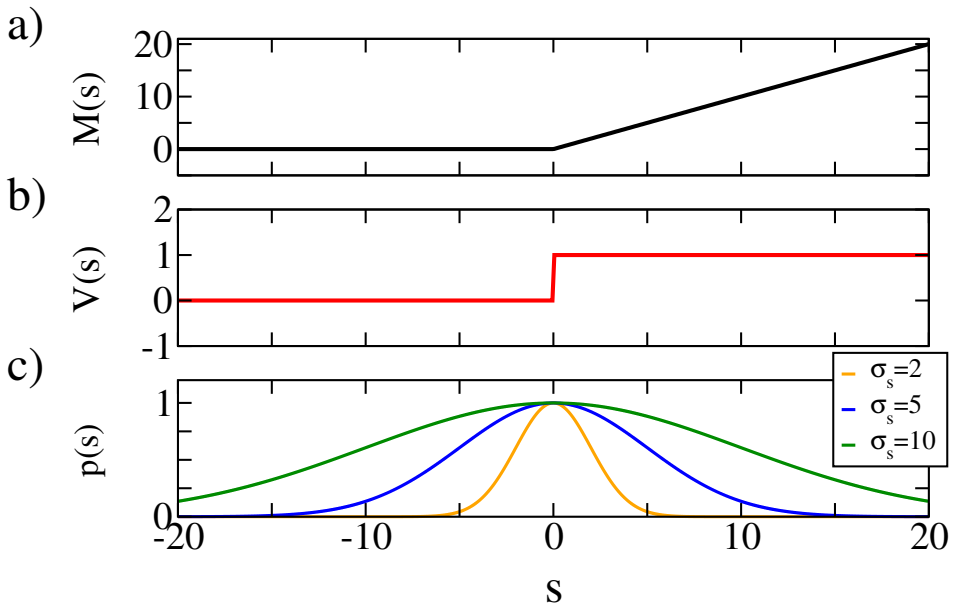

Figure 3.10: Mean, variance, and signal distributions for the piecewise linear Gaussian model. a) Function $M(s)$ vs signal-value; b) Function $V(s)$ vs signal-value; c) Signal distributions, $p(s)$, as functions of the signal-value for different σ_s rescaled such that $p(s=0)=1$.

with $M(s)$ and $V(s)$ from Eq. (3.110) one finds

$$\rho^2_{x^2,s} = \frac{\left(1+\frac{1}{2}\frac{1}{\sigma_s^2}\right)^2}{\pi\left(\frac{5}{8}+\frac{1}{\sigma_s^2}+\frac{1}{\sigma_s^4}\right)} \tag{3.115}$$

and the squared correlation coefficient between x and x^2, Eq. (3.69), reads

$$\rho^2_{x^2,x} = \frac{\frac{9}{8\pi}\left(1+\frac{4}{3}\frac{1}{\sigma_s^2}\right)^2}{\left(\frac{5}{8}+\frac{1}{\sigma_s^2}+\frac{1}{\sigma_s^4}\right)\left(1-\frac{1}{\pi}+2\frac{1}{\sigma_s^2}\right)} \,. \tag{3.116}$$

The linear lower bound, which is obtained by inserting Eq. (3.112) into Eq. (3.21) and the quadratic lower bound, which is obtained by inserting Eq. (3.112), Eq. (3.115) and Eq. (3.116) into Eq. (3.33) are both plotted in Fig. 3.11.

In the limit of $\sigma_s \to \infty$ the nonlinear coefficient in Eq. (3.115) reduces to

$$\lim_{\sigma_s\to\infty} \rho^2_{x^2,s} = \frac{8}{5\pi} \,, \tag{3.117}$$

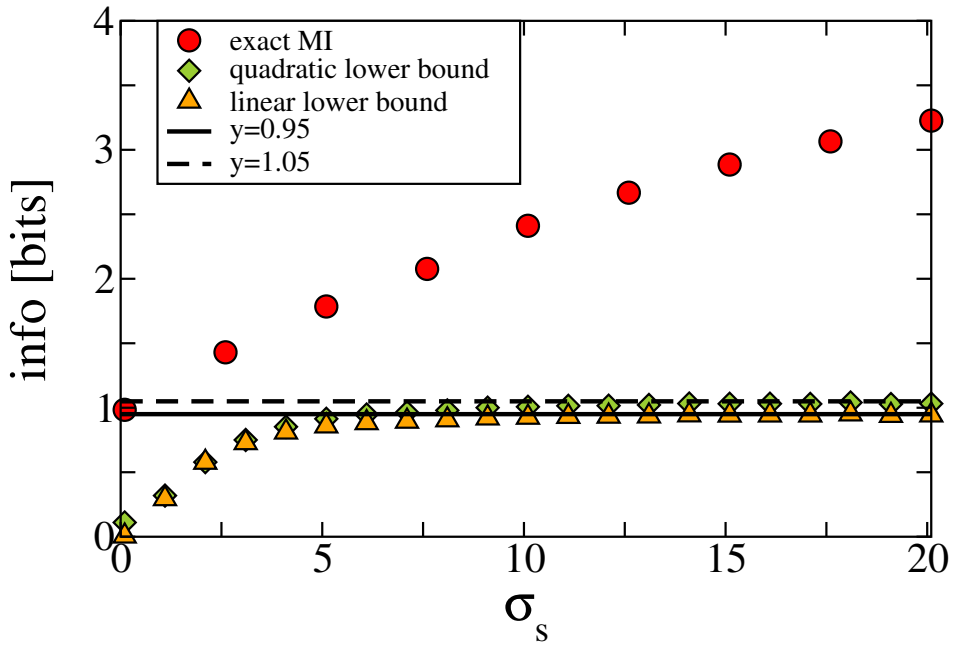

Figure 3.11: Mutual information (red circles), linear lower bound (orange triangles) and the quadratic lower bound (green diamonds) for the piecewise linear Gaussian model. The limits for $\sigma_s \to \infty$ of the linear lower bound (solid line) and of the quadratic lower bound (dashed line) are obtained from Eq. (3.114) and Eq. (3.119), respectively.

and the coefficient in Eq. (3.116) simplifies to

$$\lim_{\sigma_s \to \infty} \rho^2_{x^2,x} = \frac{9}{5\,(\pi - 1)} \,. \tag{3.118}$$

The limits in Eq. (3.117) and Eq. (3.118) lead to a finite value of the quadratic lower bound, Eq. (3.33), in the limit of strong signals

$$\lim_{\sigma_s \to \infty} MI^{quad}_{lb} \approx 1.05 \text{ bits} \,. \tag{3.119}$$

For $\sigma_s \leq 5$ the quadratic lower bound (green diamonds) and the linear lower bound (orange triangles) both increase with increasing σ_s and are hardly distinguishable from each other. For $\sigma_s > 10$ the linear lower bound saturates at 0.95 bits (solid line) and the quadratic lower bound saturates to 1.05 bits (dashed line). These are exactly the values which are predicted for the linear and the quadratic lower bounds in the limit of $\sigma_s \to \infty$ by Eq. (3.114) and Eq. (3.119), respectively. The exact mutual information is computed via numerically evaluating Eq. (3.70) with $M(s)$ and $V(s)$ from Eq. (3.110) and is indicated by the red circles in Fig. 3.11. For $\sigma_s \to 0$, the mutual information approaches the value of 1 bit. This is due to the fact that a negative signal always

leads to an output which is exactly zero, while a positive signal (even if it is very weak) will with a high probability lead to an output which is nonzero due to the intrinsic noise. Consequently, even for vanishingly small signal strength, the output allows to asses whether the signal has positive or negative sign, which corresponds to 1 bit of information. For increasing σ_s the mutual information increases monotonically without any saturation. Note, that the upper bound is ill-defined and cannot be plotted for the piecewise linear Gaussian model because the logarithm in Eq. (3.41) diverges when $V(s)$ equals to zero for a non-vanishing range of singal-values.

In conclusion, we find that for the piecewise linear Gaussian model the mutual information grows without bounds. Furthermore, in the limit of large σ_s both lower bounds converge to slightly different limits that are both different from zero. Interestingly, for the piecewise linear Gaussian model none of the two lower bounds comprises a good estimate for the exact mutual information. It seems that neither a linear reconstruction nor a reconstruction with a smooth (quadratic) nonlinearity are suitable to extract information from a highly nonlinear model as the piecewise linear Gaussian model.

3.5 Nonlinear decoding of a static signal variable from the neural spike count of the LIF and NaK models

In this section we will study the mutual information between a static input signal s and the spike count N of the LIF model and the NaK model, respectively. In particular, we will study the conditions for which the mutual information can be successfully estimated by the linear lower bound, the quadratic lower bound or the upper bound for the mutual information. Furthermore, we will investigate under which conditions a quadratic signal reconstruction performs better than a linear reconstruction. This question will be studied by considering the linear and the quadratic lower bounds and by specifying the conditions for which these two bounds significantly differ from each other.

Throughout this section we will employ $\varepsilon = 1$ in Eq. (1.33) and Eq. (1.43) but consider a varying standard deviation σ_s of the Gaussian signal distribution. The mean of the signal is assumed to be zero. The mutual information will be computed numerically via Eq. (3.52). The linear and quadratic lower bounds will be computed by numerically evaluating the correlation coefficients $\rho_{s,N}$, ρ_{s,N^2} and ρ_{N,N^2} and employing Eq. (3.21) and Eq. (3.33), respectively. The upper bound will be computed via numerically measuring the signal-dependent spike count variance and by employing Eq. (3.41).

We will now shortly discuss how the Gaussian model from the previous section can be related to the neural spike count. First, we note that the spike trains of the LIF model and the NaK model are renewal spike trains, i.e. subsequent intervals between spikes are independent of each other. The spike count N is determined by counting the spikes, which are generated by the neuron within a time window T. The models are subject to intrinsic noise, such that the spike count for a given input signal will exhibit trial to trial variability. If the time window T is large enough such that sufficiently many spikes occur within this time window, then the spike count distribution for a fixed value of the input signal is approximately Gaussian [94]. Consequently, we can use for the

signal-dependent spike count N a Gaussian model,

$$x = M(s) + \sqrt{V(s)}\,\zeta\,, \tag{3.120}$$

where ζ is a Gaussian noise with zero mean and unit variance. The function $M(s)$ in Eq. (3.120) is given by the signal-dependent mean spike count

$$M(s) = \langle N(s) \rangle_\xi \tag{3.121}$$

and the function $V(s)$ in Eq. (3.120) is given by the signal-dependent spike count variance

$$V(s) = \left\langle \Delta N^2(s) \right\rangle_\xi\,. \tag{3.122}$$

We can now use the results from the previous section to compute the mutual information and its lower and upper bounds for a signal s and the output of the Gaussian model in Eq. (3.120). The mutual information is obtained from Eq. (3.70). The linear and quadratic lower bounds are obtained by inserting Eqs. (3.67-3.69) into Eq. (3.21) and Eq. (3.33). The upper bound is computed by inserting $V(s)$ into Eq. (3.41).

LIF model

For the LIF model we can obtain an analytical prediction of the function $M(s)$. First we use that $M(s)$ can be expressed in terms of the unperturbed firing rate r_0:

$$M(s) = T\, r_0(\mu')\Big|_{\mu'=\mu+s}\,. \tag{3.123}$$

Originally, the function $r_0(\mu)$ is the unperturbed firing rate, i.e. it is the firing rate in absence of a signal, and is a function of the mean current μ. However, because here we consider a static input signal, the function r_0 can be used to describe the signal-dependent rate. The notation with the vertical line in Eq. (3.123) means that the argument of the function r_0 has to be replaced by $(\mu + s)$. For the LIF, $r_0(\mu)$ can be computed via Eq. (1.40).

For the variance, we can use Eq. (1.5) and express the variance in terms of the Fano factor and the unperturbed firing rate as follows

$$\left\langle \Delta N^2(s) \right\rangle_\xi = F(s)\,\langle N \rangle_\xi = T r_0(s) F(s)\,. \tag{3.124}$$

The Fano factor in the limit of large observation times T can be related to the low-frequency limit of the spike-train power spectrum [95]

$$\lim_{T\to\infty} F(s) = \lim_{\omega\to 0} \frac{S_{xx}(\omega, s)}{r_0(s)}\,. \tag{3.125}$$

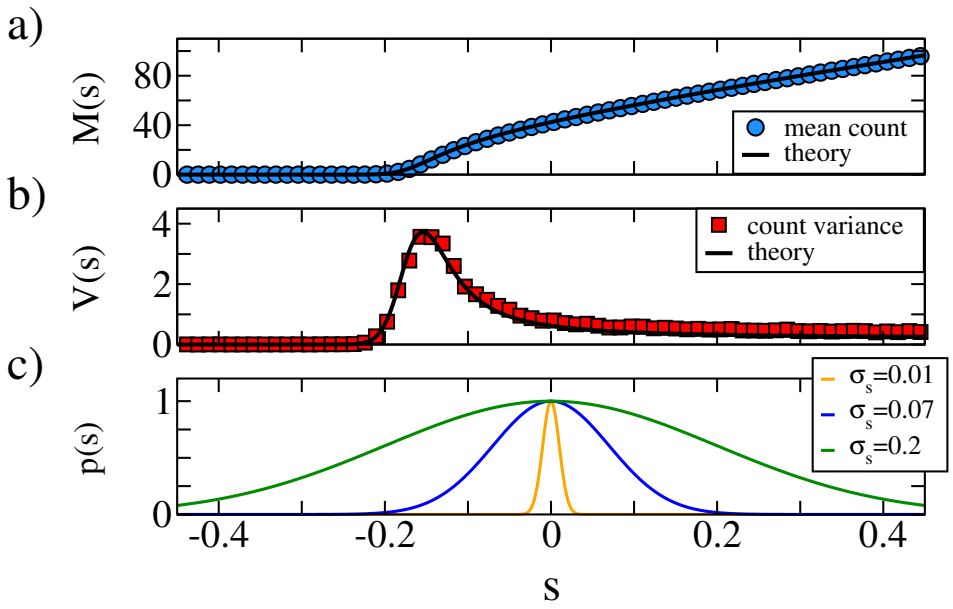

Figure 3.12: Nonlinear signal dependence of the mean count and the count variance for the LIF neuron model. a) Mean count vs value of the static input signal; b) Count variance vs value of the static input signal; c) Signal distributions for different σ_s rescaled such that $p(s=0)=1$. The signal distributions indicate the range of values which are predominantly assumed by the stochastic signal. This in turn determines which parts of the functions $M(s)$ and $V(s)$ are predominantly sampled by the signal. Model parameters are $D=0.001$, $\mu=1.1$. The time window is $T=100$.

Finally, we obtain

$$V(s) = T\, r_0(\mu')\Big|_{\mu'=\mu+s} \lim_{\omega\to 0} \frac{S_{xx}(\omega,\mu')\Big|_{\mu'=\mu+s}}{r_0(\mu')\Big|_{\mu'=\mu+s}} \,. \tag{3.126}$$

For the LIF model, the power spectrum S_{xx} in Eq. (3.126) can be computed via Eq. (1.41).

In Fig. 3.12 we compare the numerically measured mean spike count and the numerically measured spike count variance with the theoretical predictions from Eq. (3.123) and Eq. (3.126). The spikes are counted within a time-window of $T=100$. For this time-window the mean spike count in the absence of a signal is $\langle N(s=0)\rangle \approx 42$. For a membrane time-constant of $\tau_m = 10$ms in Eq. (1.33) the time interval which we consider here would correspond to a duration of 1 second. As is expected for large T, the agreement between the Gaussian model (black lines) and the LIF model (symbols) is very good. For signal values $s<-0.2$ the mean count decays to zero, while for signal values $s>0$ the mean count grows linearly with s. For small negative signal values, the mean count exhibits a nonlinear signal dependence. The count variance decays to zero

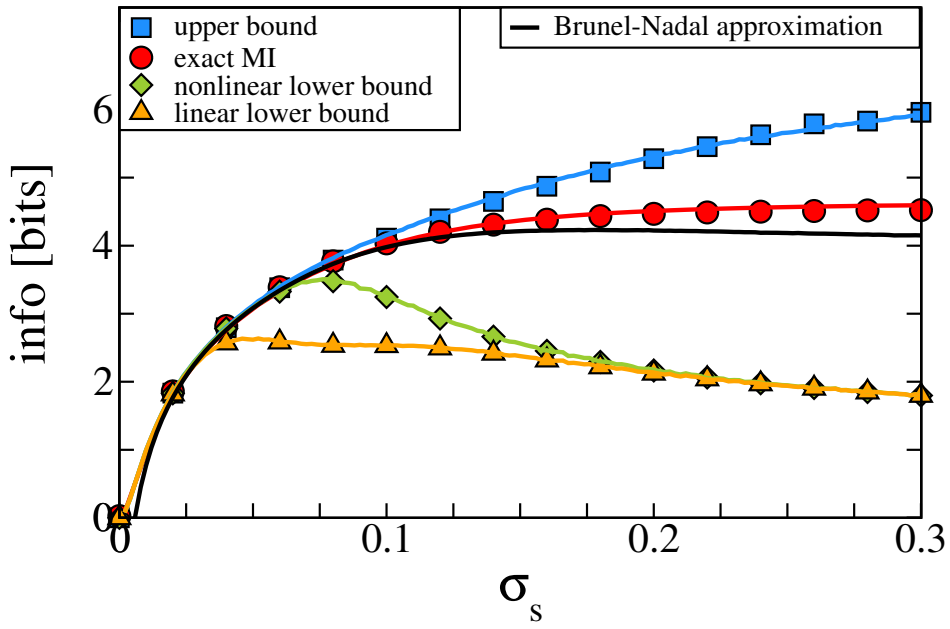

Figure 3.13: Mutual information (red circles), linear lower bound (orange triangles) and the quadratic lower bound (green diamonds) for the spike count of the LIF model for $\mu = 1.1$ vs σ_s. The colored lines represent the theoretical prediction of the mutual information (red line), the linear lower bound (orange line), the quadratic lower bound (green line) and the upper bound (blue line) from the Gaussian model. The black line represents the Brunel-Nadal theory from Eq. (3.128). The time window within which we counted the spikes is $T = 100$. The strength of the intrinsic noise is $D = 0.001$.

for small signal values and remains essentially constant for $s > 0.2$. For $-0.2 \leq s \leq 0.2$ the variance exhibits a strong nonlinear signal dependence. In particular, the variance exhibits a strong peak at around $s = -0.15$, which is reminiscent of the giant acceleration of diffusion in a tilted periodic potential [96, 97]. For a signal value of $s = -0.15$ the mean current $\mu + s$ of the LIF neuron is slightly below the threshold. Without noise ($D = 0$ in Eq. (1.33)), the LIF voltage would remain below the threshold for all times and the model would not fire at all. However, for an ensemble of LIF neurons with weak intrinsic noise, some trajectories are kicked over the threshold, while some trajectories still remain below the threshold for a long time. This separation of the ensemble leads to the enhancement of the count variance for signal values around $s = -0.15$. We expect, that the peak of the count variance shifts towards $s = v_T - \mu$ for decreasing noise strength, which with $\mu = 1.1$ and $v_T = 1$ would correspond to $s = -0.1$. For increasing noise strength, we expect the peak for the count variance to broaden and to shift towards smaller values of s.

The exact mutual information, the linear lower bound, the quadratic lower bound and the upper bound are plotted vs. σ_s in Fig. 3.13. The mutual information (red

circles) increases monotonically for increasing σ_s in a similar manner as for the linear or the weakly nonlinear Gaussian model from the previous section. The linear lower bound (orange triangles) increases rapidly with increasing signal strength for small σ_s, indicating an increase of the linear correlation between input and output. Around $\sigma_s = 0.05$ the linear lower bound reaches a peak of approximately 2.5 bits and decreases again for stronger signals. The quadratic lower bound (green diamonds) coincides with the linear lower bound for small σ_s. However, for signals of intermediate strength ($0.03 < \sigma_s < 0.13$) the quadratic lower bound significantly exceeds the linear lower bound. The largest improvement is found for $\sigma_s \approx 0.075$ where the quadratic reconstruction extracts around 40% more information (≈ 3.5 bits) than the linear reconstruction (≈ 2.5 bits). Furthermore, for $0.03 < \sigma_s < 0.07$ the quadratic lower bound constitutes a much better estimate for the exact mutual information than the linear lower bound. The upper bound (blue squares) matches the mutual information closely for $\sigma_s < 0.1$ but overestimates the mutual information significantly for $\sigma_s > 0.1$. If we did not measure the mutual information in Fig. 3.13, we could still infer its value at least for weak signals from the quadratic lower bound and the upper bound, the values of which match closely for $\sigma_s < 0.07$. If we instead used the linear lower bound and the upper bound, we could estimate the mutual information with a high accuracy only for $\sigma_s < 0.03$. For strong signals, the quadratic lower bound significantly underestimates the mutual information and the difference between the quadratic lower bound and the linear lower bound decreases.

The different behavior of the lower bounds for small, intermediate and large σ_s can be understood by considering the probability density of the signal for these three regimes in Fig. 3.12c in comparison with the functions $M(s)$ and $V(s)$. For $\sigma_s = 0.01$, the probability distribution of the signal is very narrow and the stochastic signal will mostly assume values which are narrowly distributed around the mean value of the distribution. Consequently, most of the signal values will fall in an interval for which the mean count in Fig. 3.12 is essentially linear and the count variance is nearly constant. The linear behavior of the mean count and the signal-independence of the count variance are the reason why the linear lower bound is very close to the exact mutual information for small σ_s in a similar manner as for the linear Gaussian channel from the previous section.

For signals of intermediate strength ($\sigma_s = 0.07$) the distribution of the signal is wider and the values which are assumed by the stochastic input are distributed over a wider range than before. In particular, the stochastic signal now also samples the weakly nonlinear part of the mean count and also the strongly signal-dependent part of the count variance ($-0.2 < s < 0$ in Fig. 3.12a,b). The weakly nonlinear signal dependence of the mean count and the strong signal dependence of the count variance are the reason why for $\sigma_s = 0.07$ the quadratic lower bound is very close to the exact mutual information while the linear lower bound strongly underestimates the mutual information. This behavior is comparable to the behavior of the weakly nonlinear Gaussian model from the previous section.

Finally, for $\sigma_s = 0.2$ the probability distribution of the signal in Fig. 3.12c is very wide and the stochastic input can attain very large values ($s > 0.2$) and very small values

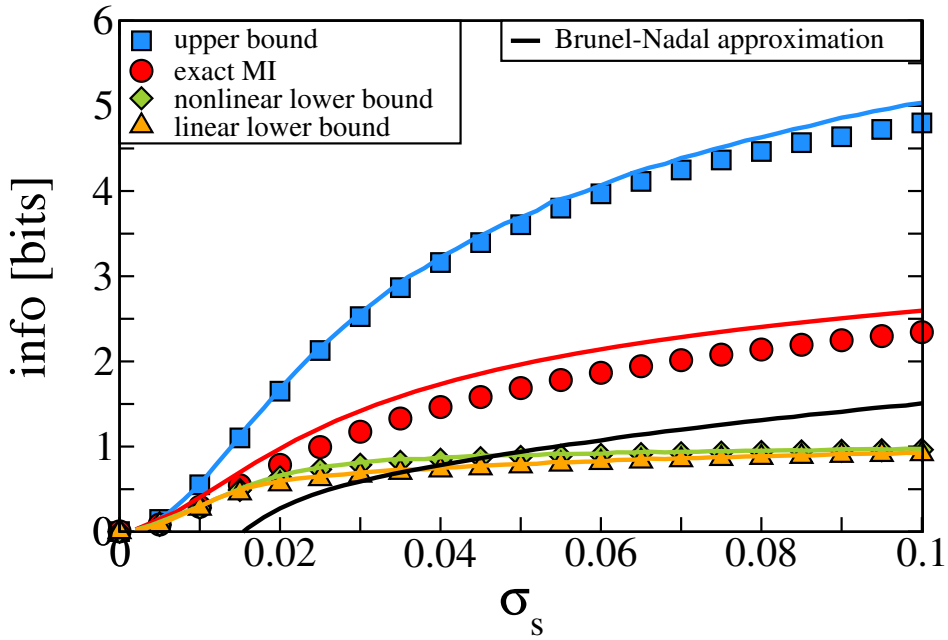

Figure 3.14: Mutual information (red circles), linear lower bound (orange triangles) and the quadratic lower bound (green diamonds) for the spike count of the LIF model for $\mu = 0.9$ vs σ_s. The colored lines represent the theoretical prediction of the mutual information (red line), the linear lower bound (orange line), the quadratic lower bound (green line) and the upper bound (blue line) from the Gaussian model. The black line represents the Brunel-Nadal theory from Eq. (3.128). For the simulations we used $D = 0.001$ and $T = 100$.

($s < -0.2$) with a relatively high probability. Because the mean count is essentially zero for $s < -0.2$ ($M(-0.2) \approx 1$ in Fig. 3.12), a lot of realizations of the signal are not transmitted at all. For large signal values ($s > 0.2$) the mean count is essentially a linear function of the signal and the count variance is almost constant. For such a piecewise linear transmission, the quadratic lower bound significantly underestimates the exact mutual information for large σ_s and almost coincides with the linear lower bound. This is the same behavior which we observed for the piecewise linear Gaussian model in the previous section.

In Fig. 3.14 we also study the subthreshold firing regime ($\mu = 0.9$). In this regime the mean spike count is in general smaller than for the suprathreshold firing regime. For a signal value $s = 0$ (which is the most probable signal value for the Gaussian distribution) the mean spice count is $\langle N \rangle_\xi \approx 1$. For such a low mean count the Gaussian assumption is of course violated. Therefore, it is very surprising that the mutual information of the Gaussian model (red line) only weakly overestimates the mutual information of the LIF model (red circles). Even more surprising is that the linear lower bound (orange line), the quadratic lower bound (green line) and the upper bound (blue line) of the Gaussian

model are perfectly aligned with the lower and upper bounds for the mutual information of the LIF model. In contrast to the suprathreshold regime, however, the quadratic lower bound is always close to the linear lower bound even for large σ_s.

We want to note that we are not the first ones to consider a Gaussian model of the spike count. A similar approach has been used by Brunel and Nadal in [98]. In this study the authors consider a static signal, which is encoded in the spike count. The authors note that for sufficiently large time windows the spike count can be approximated by a Gaussian variable, similarly to what we did in Eq. (3.120). In their paper, Brunel and Nadal argue that in the limit $V(s) \to 0$ the mutual information can be approximated by

$$MI_{\mathrm{BN}} = H_{s_0} + \frac{1}{2} \int_{-\infty}^{\infty} ds\, p_s(s) \log_2 \left(\frac{M'(s)^2}{V(s)} \right) - H_\zeta \,, \tag{3.127}$$

where H_{s_0} is the entropy of a variable s_0 which has the same distribution as the signal s but has unit variance, where H_ζ is the entropy of the noise variable, and where $M'(s)$ is the derivative of M with respect to s. For a Gaussian distributed signal and a Gaussian distributed noise, one finds that $H_{s_0} = H_\zeta$ and Eq. (3.127) can be simplified to

$$MI_{\mathrm{BN}} = \frac{1}{2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} ds\, e^{-\frac{s^2}{2}} \log_2 \left(\frac{\sigma_s^2 M'(s\,\sigma_s)^2}{V(s\,\sigma_s)} \right) \,. \tag{3.128}$$

For general nonlinear $M(s)$ and $V(s)$ the above equation cannot be simplified further and the integral has to be evaluated numerically. However, in contrast to the expression for the mutual information of the Gaussian model, Eq. (3.70), where two integrals have to be evaluated numerically, Eq. (3.128) requires the evaluation of only one integral.

We plot the Brunel-Nadal theory from Eq. (3.128) in Fig. 3.13 and Fig. 3.14 and compare it with the mutual information which we obtained for the LIF model and for the corresponding Gaussian model. In the suprathreshold regime in Fig. 3.13, the Brunel-Nadal theory (black line) provides a good estimate for the exact mutual information for $0.01 < \sigma_s < 0.15$ and shows only slight deviations for $\sigma_s < 0.01$ and $\sigma_s > 0.15$. In the subthreshold regime in Fig. 3.14, however, the Brunel-Nadal theory significantly underestimates the mutual information. In particular for small σ_s the Brunel-Nadal theory fails to correctly estimate the mutual information. For $\sigma_s < V(s\,\sigma_s)/M'(s\,\sigma_s)^2$ the argument of the logarithm in Eq. (3.128) becomes smaller than one such that MI_{BN} can even become negative for small standard deviations of the signal.

In conclusion, we find that the Brunel-Nadal theory provides a good estimate of the mutual information in the suprathreshold firing regime of the LIF model. However, in addition to the Gaussianity of the output the theory also relies on the smallness of the output variability. This additional assumption reduces the range of firing regimes and signal standard deviations for which the theory can provide a good estimate of the mutual information. The integral equation for the mutual information for the Gaussian model requires the numerical evaluation of two integrals but can be applied in a wider range of firing regimes and for a wider range of signal standard deviations.

In conclusion, for the quadratic lower bound in the suprathreshold regime we found that it strongly exceeds the linear lower bound for intermediate values of σ_s. At the same time it provides a good approximation for the exact mutual information. Of course, the quadratic lower bound drastically underestimates the mutual information for large σ_s. Note, however, that for $\sigma_s = 0.21$ already approximately 16% of the signal values are not transmitted because the count is essentially zero for $s < -0.21$. If only signals are considered which do not completely silence the neuron but merely modulate the mean count then the quadratic lower bound could prove to be a helpful tool for the estimation of the mutual information.

We can also interpret our results not only from an information theoretic perspective but also from a decoding perspective. The strong deviation of the quadratic lower bound from the linear lower bound indicates that the quadratic reconstruction, Eq. (3.22), outperforms the linear reconstruction from Eq. (3.18). Additionally, the good agreement of the quadratic lower bound and the exact mutual information indicates that for signals of intermediate strength the quadratic reconstruction is optimal in the sense that it recovers almost all of the information that is present in the output. Because the mutual information of the LIF model can be successfully described by the Gaussian model, we can infer that the main reason for the nonlinear decoding in the LIF model stems from the nonlinear signal dependencies of the mean count and the count variance.

NaK model

In order to test the robustness of the results which we obtained for the LIF neuron model, we will now consider the decoding of a static signal from the spike count of the NaK model with ion channel noise. In Fig. 3.15a,b we plot the numerically measured mean count and count variance vs signal value s. We consider a time window $T = 1\,\text{second}$ for which in the absence of a signal the mean number of spikes is $\langle N(s=0)\rangle \approx 100$. We find that the mean count and the count variance of the NaK model both exhibit a nonlinear signal dependence, which is similar to what we observed for the LIF model.

The mutual information, the linear lower bound, the quadratic lower bound and the upper bound are plotted in Fig. 3.16. The mutual information (red circles) rises monotonically for increasing signal strength. The upper bound (blue squares) matches the mutual information closely for $\sigma_s < 0.4\mu\text{A/cm}^2$ but overestimates the mutual information for stronger signals. The linear lower bound (orange triangles) and the quadratic lower bound (green diamonds) have very similar values for $\sigma_s \leq 0.3\mu\text{A/cm}^2$ but for $\sigma_s > 0.3\mu\text{A/cm}^2$ the quadratic lower bound significantly exceeds the linear lower bound. Furthermore, for $0.3\mu\text{A/cm}^2 < \sigma_s < 0.5\mu\text{A/cm}^2$ the quadratic lower bound constitutes a much better estimate for the exact mutual information than the linear lower bound. Note, that the mutual information, the linear lower bound, the quadratic lower bound and the upper bound for the Gaussian model (red, orange, green and blue lines, respectively) closely match the corresponding measures for the NaK model. Overall, the mutual information and its lower and upper bounds exhibit a very similar behavior to what we observed for the LIF model in the suprathreshold firing regime. It is interesting, that nonlinear decoding of input signals can not only be observed for the LIF model

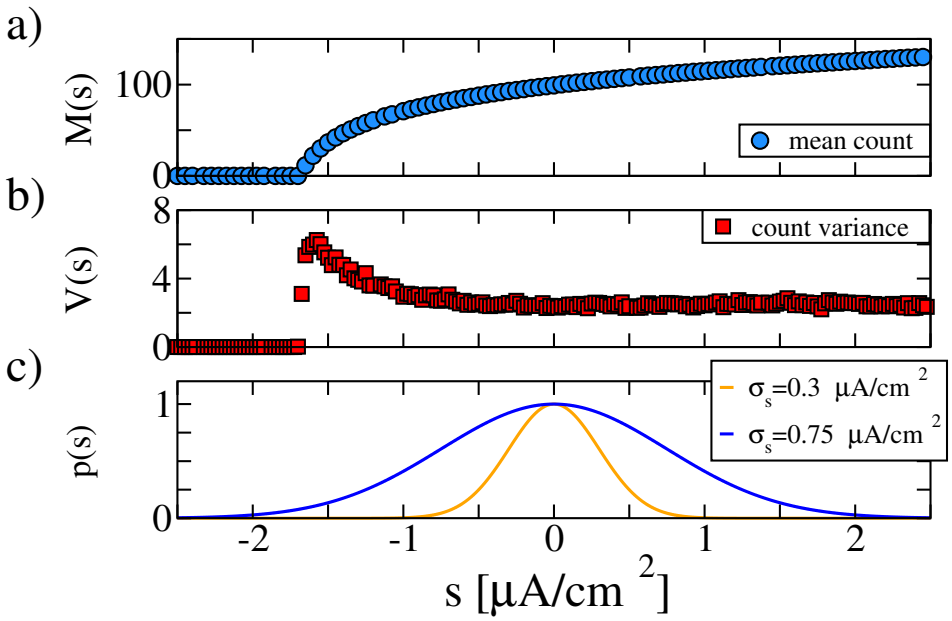

Figure 3.15: Nonlinear signal dependence of the mean count and the count variance for the NaK neuron model. a) Mean count vs value of the static input signal; b) Count variance vs value of the static input signal; c) Signal distributions for different σ_s rescaled such that $p(s = 0) = 1$. The signal distributions indicate the range of values which are predominantly assumed by the stochastic signal. This in turn determines which parts of the functions $M(s)$ and $V(s)$ are predominantly sampled by the signal. The time window within which the spikes were counted was $T = 1$ second. The static input current is $I_0 = 6\mu\text{A/cm}^2$.

with dynamic Gaussian noise but also for a neuron model with a realistic noise source (discrete channel noise).

3.6 Summary

In this chapter, we first derived a quadratic lower bound for the mutual information between a static input signal and a static output variable for a general signal processing system. We also derived an upper bound for the mutual information which is valid for systems for which the output variability (for fixed signal) is Gaussian distributed. We then discussed the difficulties that arise for the numerical measurement of the exact mutual information between a discretized input signal and a discretized output variable and presented a numerical routine which can be used for general signal processing systems. The qualitative behavior of the different information measures was studied for a simple Gaussian model with different choices of the signal dependency of the mean and the variance of the output. For a signal-independent variance and a linear signal dependence of the mean, the linear lower bound and the quadratic lower bound yielded

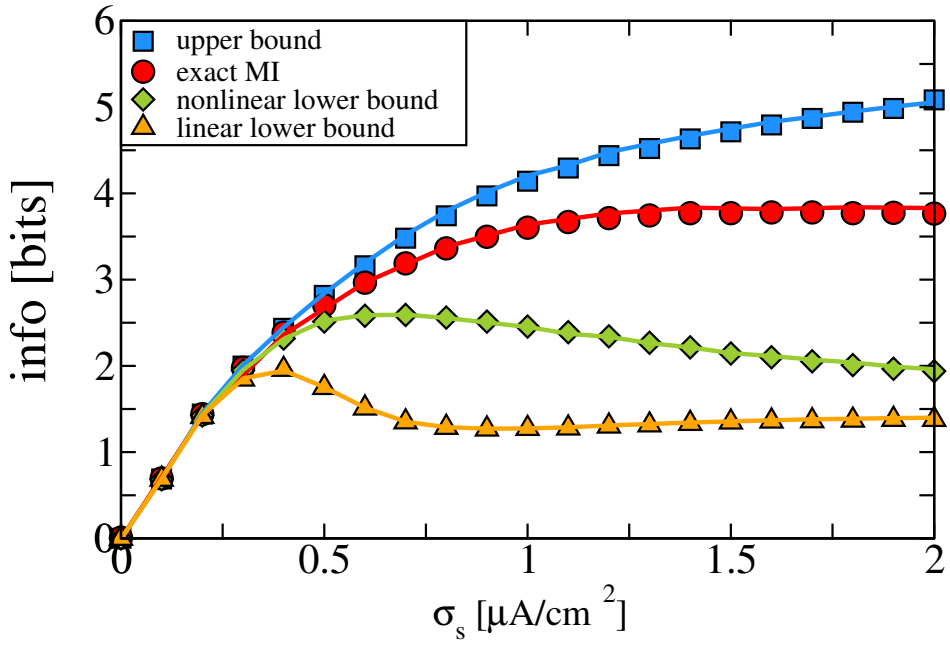

Figure 3.16: Mutual information (red circles), linear lower bound (orange triangles) and the quadratic lower bound (green diamonds) for the spike count of the NaK model for vs σ_s. The solid lines represent the theoretical prediction of the mutual information (red line), the linear lower bound (orange line), the quadratic lower bound (green line) and the upper bound (blue line) from the Gaussian model. We used $T = 1$ second and $I_0 = 6\mu\text{A/cm}^2$.

identical results. From the perspective of signal decoding this means that for a linear system the quadratic reconstruction does not provide any improvement as compared to the linear reconstruction. However, the linear lower bound and the quadratic lower bound exhibited pronounced differences for a variant of the Gaussian model with a nonlinear signal dependence of the mean and for another variant of the model in which only the variance was signal dependent. Finally, we quantified the information which the spike count of the LIF model carries about a static input signal. We found that the linear lower bound and the quadratic lower bound coincide for weak signals but that the quadratic reconstruction outperforms the linear reconstruction for intermediate signal strengths. Furthermore, the quadratic lower bound provided a good estimate of the true mutual information for a larger range of signal strengths than the linear lower bound. The results which we observed for the LIF were also replicated in the NaK model, a two dimensional model with discrete channel noise. The similarity of the results for the LIF model and the NaK model suggests that our results for the nonlinear decoding of information are not model specific but are of a more general nature.

The main result of this chapter is the derivation of the quadratic lower bound in section 3.1. Our result is expressed in terms of linear and nonlinear correlation coefficients

between input and output and can be computed for any signal processing system that encodes an input variable into an output variable. The quadratic lower bound does not only constitute a simple estimate for the mutual information but can also be interpreted from a decoding perspective: It corresponds to the minimal information that is recovered from the output via a quadratic reconstruction. A deviation of the quadratic lower bound from the linear lower bound allows to determine regimes for which the optimal decoding of a signal (and, consequently, also the encoding of information) is nonlinear. This is particularly useful in cases in which the exact mutual information is neither computable analytically nor numerically as, for example, for biological experiments for which only a limited amount of data is available.

4 Discussion

Abstract | The main results of this thesis are summarized, the implications of the findings are discussed and an outlook on possible continuations of the work is presented.

In this thesis we investigated the processing of signals by excitable nerve cells. In general, these signals are processed in a nonlinear fashion, the exact mathematical description of which is in general difficult. For weak signals, however, it is possible to describe the signal processing of neurons by two well-established linear theories.

One of these theories ist the linear response theory [35] which describes modulations of the neural firing rate by time-dependent signals [38]. The linear response of neurons has been measured in experiments [11, 12, 39] and has been studied analytically for mathematical models of the neural activity [40, 41, 42]. Furthermore, the linear response theory has not only been used for the prediction of the time-dependent firing rate for a given input signal but has also been employed for the investigation of basic properties of the neural code, such as the beneficial influence of neural noise on signal transmission [41, 42, 83, 99, 100], the stability of the activity of recurrent neural networks [45], and the speed of information transmission in neural systems [43, 44]. However, the linearity of the neural response is ensured only for weak signals and extensions of the linear response theory are required in order to describe the response of neurons for a broad range of signal strengths.

Another well-established linear theory which is often used for the investigation of signal processing in neural systems is the linear decoding theory [38]. In contrast to the linear response theory, which is used to predict the neural response to a given signal, the linear decoding theory is used to reconstruct a signal from the activity of a neuron that was driven by that signal. In particular, the linear decoding theory comprises a linear filter operation the kernel of which is optimized such that the error between the reconstructed signal and the actual signal is minimized. Such filters have been computed for biological neurons and have been shown to reconstruct input signals with a high accuracy [12, 101, 102]. The information that is retrieved from the neural output by an optimal linear filter can be related to the mutual information by means of the well-known linear lower bound [38, 63]. The mutual information allows to quantify the nonlinear

dependence between an input signal and the resulting neural activity and is a valuable tool for the investigation of signal processing in neural systems [62, 103]. The problem with the mutual information is that its computation is rather involved and requires a lot of data [61, 66], which is not always available in the required amount. Therefore, the linear lower bound, which can be computed from simple statistics of the input and output variables, has been widely employed as an estimate for the mutual information [38, 62]. However, the approximation is tight only for weak signals and has been shown to significantly underestimate the mutual information in the case of strong stimulation [65, 104] or if the neuron exhibits bursting activity [64].

In chapter 2 we considered a Volterra expansion of the time-dependent firing rate and extended the linear response theory by incorporating second-order contributions of the input signal. For the leaky integrate-and-fire model we analytically derived the second-order response function and investigated the implications of the resulting weakly nonlinear theory. This theory allows to predict the time-dependent firing rate in response to general time-dependent signals, which is not only a single-neuron statistics which is often measured in experiments, but can also be regarded as the instantaneous population rate of a group of uncoupled neurons subject to the same external drive.

In neuroscience, the second-order response of nerve cells has been measured experimentally [50, 51, 52, 53] and has also been calculated for a Poisson neuron model [55]. However, the nonlinearity of the response in the latter study resulted from an absolute refractory time of the firing rate and not from the intrinsic dynamics of the model. For the leaky integrate-and-fire model the second-order response function was already considered in the weakly nonlinear analysis of the stability of a neural network [54] but the mathematical theory did not include the complete second-order response and missed some of the second-order contributions of the signal to the rate modulation. Furthermore, no significant nonlinear effects of the rate modulation were reported and no systematic analysis of the weakly nonlinear regime was performed.

In this work, we presented the complete second-order response of the LIF model, which does not result from an absolute refractory time but from the intrinsic dynamics of the model. For cosine signals, the weakly nonlinear theory reveales several nonlinear effects: i) The excitation of higher harmonics; ii) a frequency-dependent change of the time-averaged firing rate; iii) a strong nonlinear interaction of multiple input signals. For stochastic time-dependent Gaussian signals of intermediate strength, the weakly nonlinear theory faithfully describes the exact time-course of the modulations of the neural firing rate.

As was already discussed above, the linear theory becomes exact in the limit of weak signals. In the limit of strong input signals, the neural response can be quantified by tools from nonlinear dynamics, e.g. by measures of mode locking and synchronization [105]. Although for general nonlinear systems some features of the transition between the limits of very weak and very strong signals can be described by the theory of stochastic synchronization [106, 107], this transition is analytically only poorly understood in the context of neural systems. In this work, we demonstrated that the weakly nonlinear regime emerges as a transition between the regime of linear encoding and a highly

nonlinear regime in which the neuron exhibits mode-locking behavior. The transition emerges not only for increasing signal amplitude, as suggested by the Volterra expansion of the firing rate, but also for decreasing strength of the intrinsic noise. This transition is consistent with the known linearizing effect of fluctuations which has been described for a simple static nonlinearity [108], for neural models [89], and also for biological neurons [109].

Beyond the analytically tractable leaky integrate-and-fire model with white Gaussian current noise, we also measured numerically the second-order response function for a biophysically more realistic spiking neuron model with channel noise. The similarity of the weakly nonlinear response for the two models suggests that the discussed features do not hinge on the particularities of the models but are of a more general nature.

We expect that the weakly nonlinear theory can be relevant for the description of neural responses to current stimulations in vitro for which channel noise is the main source of variability. Furthermore, the weakly nonlinear response theory may also be relevant for neural networks with low background activity [110] and in cases in which the network activity exhibits one or multiple global oscillations [111].

In chapter 3 we investigated the nonlinear reconstruction of input signals and the nonlinear estimation of the information content of the neural spike count. For the special case of a static input signal we derived a quadratic lower bound, which can be used as a simple estimate of the mutual information. Note that the linear lower bound can be related not only to the mutual information but under certain assumptions also to the Fisher information and to the probability of a successful signal detection [98, 112]. Similar arguments can also be applied to the quadratic lower bound, although in this thesis we only focused on the mutual information. Our result for the quadratic lower bound is expressed in terms of linear and nonlinear correlation coefficients which can be computed for any signal-processing system with a time-independent input and output. From a decoding perspective, the quadratic lower bound corresponds to the minimal information that is recovered from the output via a quadratic reconstruction. For a Gaussian model we demonstrated that the quadratic signal reconstruction outperforms the linear theory if the mean of the output variable exhibits a nonlinear signal dependence. More interestingly, however, the quadratic lower bound can be used as an estimate of the mutual information for systems which exhibit a signal-dependent variance and for which the linear lower bound is always zero. For two different neuron models, we demonstrated that a quadratic reconstruction of the input signal can outperform the linear reconstruction significantly. This finding implies that higher-order correlations between signal and spike count carry information which is not accounted for by the linear correlations. Naively, one may think that the significance of the quadratic decoding follows automatically from the significance of the weakly nonlinear response, which was demonstrated in chapter 2. This is, however, not the case because the nature of the expansion involved in the derivation of the weakly nonlinear response theory is very different from the expansion which leads to the quadratic lower bound. The Kernels of the Volterra expansion of the firing rate are uniquely determined by the dynamics of the model and are, in particular, independent from the applied signal. The kernels of the quadratic signal

reconstruction, however, are determined from the combined statistics of the output and the input variables and are optimized such that they minimize the average reconstruction error. Therefore, many examples exist for which the neural response is nonlinear but the signal is still successfully reconstructed by a linear reconstruction (cf. [56] and references therein). For the estimation of the mutual information between signal and spike count we find that the linear and the quadratic lower bounds tightly approximate the mutual information for weak signals. For stronger signals, however, the quadratic lower bound provides an accurate estimate of the mutual information for a larger range of signal strengths than the linear lower bound. This suggests that the quadratic lower bound may also be relevant for neural systems for which the linear lower bound has been shown to significantly underestimate the mutual information [65, 104] and it would be interesting to test wether the quadratic lower bound can provide a better estimate for the mutual information in these cases. Furthermore, the quadratic lower bound may be relevant for neural systems which exhibit a signal-dependent variability [26, 27].

The results of this thesis demonstrate how existing linear theories can be extended to capture nonlinear contributions of the signal to the neural response or to incorporate nonlinear correlations for the estimation of the transmitted information. More importantly, however, our analysis demonstrated that these extensions do not only provide a correction to the existing linear theories but can also account for qualitatively novel effects which are completely missed by the linear theories. These effects include, for example, the excitation of harmonic oscillations in the neural firing rate or the estimation of information for systems with a signal-dependent output variance.

One obvious possible extension of the results from chapter 3 would be the derivation of a quadratic lower bound for time-dependent signals wich are encoded in neural spike trains. For example, let us consider a Gaussian signal the power of which is limited to one Fourier mode at frequency ω. Then it should be possible to express the reconstruction of the Fourier mode of the signal as

$$\tilde{s}(\omega)^{quad}_{rec,1} = \tilde{h}_1(\omega)\tilde{x}(\omega) + \tilde{h}_2(\omega)\tilde{x}(2\omega)\tilde{x}(2\omega) \; . \tag{4.1}$$

If the second term in the above equation is omitted, then the well-known linear reconstruction is recovered. The reconstruction in Eq. (4.1) does not exactly constitute a Volterra expansion of the reconstructed signal. But as we discussed earlier this is not a problem because the kernels are not determined by the dynamical properties of the neuron but can be considered as free parameters of an optimization routine. If we properly optimize the reconstruction in Eq. (4.1), the coefficient $\tilde{h}_2$ will either be zero (in which case the linear reconstruction is recovered) or the coefficient will be nonzero, in which case the reconstruction in Eq. (4.1) should perform better than the linear reconstruction. In this case it should be possible to express the quadratic lower bound in terms of higher-order correlation coefficients between the Fourier mode of the signal and the Fourier modes of the output which are also referred to as bicoherencies [113]. Indeed, such bicoherencies have been measured in experiments and it has been shown that they can capture nonlinear contributions of strong input signals to the neural response [104]. The problem with the reconstruction in Eq. (4.1) is that its second-order term is

not uniquely defined. In Eq. (4.1), for example, we included the Fourier mode of the spike train at 2ω. The idea was that these modes could help to decode higher-harmonic oscillations which might be excited by the signal. Other possible choices could be

$$\tilde{s}(\omega)^{quad}_{rec,2} = \tilde{h}_1(\omega)\tilde{x}(\omega) + \tilde{h}_2(\omega)\tilde{x}(\omega/2)\tilde{x}(\omega/2) \tag{4.2}$$

or

$$\tilde{s}(\omega)^{quad}_{rec,3} = \tilde{h}_1(\omega)\tilde{x}(\omega) + \tilde{h}_2(\omega)\tilde{x}(\omega)\tilde{x}(\omega) \ . \tag{4.3}$$

Consequently, the major difficulty of the extension of the quadratic lower bound to time-dependent signals is to smartly choose the second-order term which will lead to a significant improvement of the quadratic reconstruction out of many possible (and possibly useless) choices.

Another possible extension of the results in chapter 3 could be the derivation of a quadratic lower bound for the mutual information between a signal and the joint activity of multiple neurons. For two neurons, one possible choice of a reconstruction of the signal could be

$$s^{quad}_{rec} = h_1 N_1 + h_2 N_2 \ , \tag{4.4}$$

where N_1 and N_2 are the spike counts of the first and second neuron, respectively. We expect that it should be possible to perform an optimization of the reconstruction in Eq. (4.4) in a similar manner as we did in chapter 3 for the quadratic reconstruction and that the resulting lower bound for the mutual information between the signal and the joint spike count of the two neurons could be expressed in terms of the correlation coefficients ρ_{s,N_1}, ρ_{s,N_2}, and ρ_{N_1,N_2}. Considering the large interest in the role of inter-neuronal correlations for signal processing in neural systems [114, 115, 116], a simple estimate of the mutual information between a signal and the joint activity of neurons could be useful. Note, however, that if the spike counts of a population are considered for time-dependent signals then the same problems are encountered as the ones which we discussed for time-dependent signals in the single-neuron case.

Possible extensions of the framework from chapter 2 could be the application of the Richardson method [117] for an efficient numerical estimation of higher-order response functions of integrate-and-fire models. Furthermore, it would be interesting to investigate the weakly nonlinear response of neural models with finite synaptic potentials [118, 119] because these models are known to exhibit a nonlinear rate response. Another possible extension our investigation of the signal dependence of the time-dependent firing rate, which is a first-order spike train statistics, could be the investigation of the nonlinear signal dependence of higher-order spike train statistics such as the power spectrum or the coherence function.

A Simulation parameters

Abstract | We list the parameters of the numerical simulations, which we used for the figures. The parameters which were already mentioned in the main text are omitted.

Throughout this work, the errors of results from numerical simulations are within symbol size, unless stated otherwise.

Figures chapter 1

Fig. 1.1	$\Delta t_{int} = 10^{-2}$ (integration time step), $\Delta t_{hist} = 10^{-2}$ (time-interval, which was used for the binning of the firing rate histogram), $\tau_r = 0$ (refractory time).
Fig. 1.2	$\Delta t_{int} = 10^{-2}$, $\Delta t_{hist} = 0.2$.

Figures chapter 2

Fig. 2.1	$\tau_r = 0$, $\Delta\omega/2\pi = 10^{-2}$ (frequency resolution of the 2-d plots); A linear interpolation between the data points was used for the 2-d plots.
Fig. 2.2	$\Delta t_{int} = 10^{-3}$, $\Delta f = 10^{-3}$ (frequency resolution of the Fourier transform), $N_{av} = 10^2$ (number of trials used for averaging), $\tau_r = 0$.
Fig. 2.3, Fig. 2.4	$\Delta t_{int} = 10^{-2}$, $\Delta t_{hist} = 10^{-2}$, $N_{av} = 10^6$, $\tau_r = 0$.

Fig. 2.5a	$\Delta t_{int} = 10^{-6}$, $N_{av} = 10^6$.
Fig. 2.5b	$\Delta t_{int} = 5 \cdot 10^{-6}$, $N_{av} = 2 \cdot 10^5$.
Fig. 2.6	$\Delta t_{int} = 10^{-3}$, $T = 10^4$ (time window); The mean rate was determined by dividing the number of total spikes by T.
Fig. 2.7	$\Delta t_{int} = 10^{-2}$, $\Delta t_{hist} = 10^{-2}$, $N_{av} = 10^5$.
Fig. 2.8	$\Delta t_{int} = 10^{-2}$, $\Delta t_{hist} = 10^{-1}$, $N_{av} = 10^6$.
Fig. 2.9, Fig. 2.10	$\Delta t_{int} = 10^{-2}$, $\Delta t_{hist} = 10^{-2}$, $N_{av} = 10^5$.
Fig. 2.11	$\Delta t_{int} = 10^{-2}$, $\Delta t_{hist} = 0.57$, $N_{av} = 5 \cdot 10^4$.
Fig. 2.12a,b	$\Delta t_{int} = 0.01$ms, $\Delta\omega/2\pi = 5$Hz, $N_{av} = 3 \cdot 10^3$.
Fig. 2.12c,d	$\Delta t_{int} = 0.01$ms, $\Delta\omega/2\pi = 10$Hz, $N_{av} = 3 \cdot 10^3$.

Figures chapter 3

Fig. 3.1, Fig. 3.2, Fig. 3.3	$s_{min} = -4,5\sigma_s$ (lower integration boundary for the numerical integration over s), $s_{max} = 4.5\sigma_s$ (upper integration boundary for the numerical integration over s), $x_{min} = -45$, $x_{max} = 45$.
Fig. 3.4, Fig. 3.6, Fig. 3.8, Fig. 3.10	All relevant parameters are listed in the text.
Fig. 3.5	Lower and upper bounds: $N_{av} = 10^4$; mutual information: $s_{min} = -3\sigma_s$, $s_{max} = 3\sigma_s$, $\Delta s_{int} = 2 \cdot 10^{-2}$ (step size of the numerical integration), $x_{min} = -350$, $x_{max} = 350$, $\Delta x_{int} = 5 \cdot 10^{-2}$ (step size of the numerical integration).
Fig. 3.7	Lower and upper bounds: $N_{av} = 10^6$; mutual information: $s_{min} = -3\sigma_s$, $s_{max} = 3\sigma_s$, $\Delta s_{int} = 10^{-2}$, $x_{min} = -300$, $x_{max} = M(s_{max}) + 3$, $\Delta x_{int} = 10^{-1}$.
Fig. 3.9	Lower and upper bounds: $N_{av} = 10^5$; mutual information: $s_{min} = -3\sigma_s$, $s_{max} = 3\sigma_s$, $\Delta s_{int} = 5 \cdot 10^{-3}$, $x_{min} = -80$, $x_{max} = 80$, $\Delta x_{int} = 5 \cdot 10^{-3}$.

Fig. 3.11 Lower and upper bounds: $N_{av} = 10^5$; mutual information: $s_{min} = -3\sigma_s$, $s_{max} = 3\sigma_s$, $\Delta s_{int} = 5 \cdot 10^{-2}$, $x_{min} = -3$, $x_{max} = 120$, $\Delta x_{int} = 10^{-2}$.

Fig. 3.12 $\Delta t_{int} = 10^{-3}$, $N_{av} = 10^3$.

Fig. 3.13 LIF: $\Delta t_{int} = 10^{-2}$; mutual information and lower bounds for the LIF model: $N_{av} = 10^5$, $s_{min} = -4.5\sigma_s$, $s_{max} = 4.5\sigma_s$, $\Delta s = 10^{-2}$; upper bound for the LIF model: $N_{av} = 10^6$, $s_{min} = -4.5\sigma_s$, $s_{max} = 4.5\sigma_s$, $\Delta s = 5 \cdot 10^{-2}$; lower and upper bounds for the Gaussian model: $N_{av} = 10^6$; mutual information for the Gaussian model: $s_{min} = -3\sigma_s$, $s_{max} = 3\sigma_s$, $\Delta s_{int} = 5 \cdot 10^{-2}$, $x_{min} = -3$, $x_{max} = M(s_{max}) + 3\sqrt{V(s_{\max})}$, $\Delta x_{int} = 5 \cdot 10^{-2}$; mutual information for the Brunel-Nadal theory: $\Delta s = 5 \cdot 10^{-3}$.

Fig. 3.14 LIF: $\Delta t_{int} = 10^{-2}$; mutual information and lower bounds for the LIF model: $N_{av} = 10^6$, $s_{min} = -3\sigma_s$, $s_{max} = 3\sigma_s$, $\Delta s = 75 \cdot 10^{-4}$; lower and upper bounds for the Gaussian model: $N_{av} = 10^5$; mutual information for the Gaussian model: $s_{min} = -3\sigma_s$, $s_{max} = 3\sigma_s$, $\Delta s_{int} = 5 \cdot 10^{-2}$, $x_{min} = -3$, $x_{max} = M(s_{max}) + 3\sqrt{V(s_{\max})}$, $\Delta x_{int} = 5 \cdot 10^{-2}$; mutual information for the Brunel-Nadal theory: $\Delta s = 5 \cdot 10^{-3}$.

Fig. 3.15 $\Delta t_{int} = 10^{-2}$, $N_{av} = 10^3$.

Fig. 3.16 NaK: $\Delta t_{int} = 10^{-2}$; $N_{av} = 10^5$, $s_{min} = -3\sigma_s$, $s_{max} = 3\sigma_s$, $\Delta s = 6 \cdot 10^{-2}$; lower and upper bounds for the Gaussian model: $N_{av} = 10^5$; mutual information for the Gaussian model: $s_{min} = -3\sigma_s$, $s_{max} = 3\sigma_s$, $\Delta s_{int} = 10^{-1}$, $x_{min} = -3$, $x_{max} = 300$, $\Delta x_{int} = 10^{-1}$.

Bibliography

[1] E. R. Kandel and L. R. Squire. Neuroscience: Breaking down scientific barriers to the study of brain and mind. *Science*, 290(5494):1113–1120, 2000.

[2] J. C. A. Read. The place of human psychophysics in modern neuroscience. *Neuroscience*, 296:116–129, 2015.

[3] N. K. Logothetis. What we can do and what we cannot do with fMRI. *Nature*, 453(7197):869–878, 2008.

[4] F. L. da Silva. EEG and MEG: Relevance to Neuroscience. *Neuron*, 80(5):1112–1128, 2013.

[5] A. R. Houweling and M. Brecht. Behavioural report of single neuron stimulation in somatosensory cortex. *Nature*, 451:65, 2008.

[6] G. B. Stanley. Reading and writing the neural code. *Nature Neuroscience*, 2013.

[7] S. Panzeri, C. D. Harvey, E. Piasini, P. E. Latham, and T. Fellin. Cracking the Neural Code for Sensory Perception by Combining Statistics, Intervention, and Behavior. *Neuron*, 93(3):491–507, 2017.

[8] W. Gerstner and R. Naud. How good are neuron models? *Science*, 2009.

[9] R. Brette. What Is the Most Realistic Single-Compartment Model of Spike Initiation? *PLoS Comput. Biol.*, 11(4):e1004114–13, 2015.

[10] E. M. Izhikevich. *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting*. The MIT Press, Cambridge, London, 2007.

[11] B. W. Knight. The relationship between the firing rate of a single neuron and the level of activity in a population of neurons. *Journal of general physiology*, 59:767, 1972.

[12] W. Bialek, F. Rieke, R. R. D. Vansteveninck, and D. Warland. Reading a neural code. *Science*, 252:1854, 1991.

[13] S. O. Voronenko and B. Lindner. Weakly nonlinear response of noisy neurons. *New Journal of Physics*, 19(3):033038, 2017.

[14] S. O. Voronenko and B. Lindner. Nonlinear information transmission by the neural spike count. *in preparation*, 2017.

[15] C. Allen and C. F. Stevens. An Evaluation of Causes for Unreliability of Synaptic Transmission. *PNAS*, 91(22):10380–10383, 1994.

[16] R. Miledi. Spontaneous synaptic potentials and quantal release of transmitter in the stellate ganglion of the squid. *The Journal of Physiology*, 192(2):379–406, 1967.

[17] J. A. White, J. T. Rubinstein, and A. R. Kay. Channel noise in neurons. *Trends Neurosci.*, page 131, 2000.

[18] A. Destexhe, M. Rudolph, and D. Paré. The high-conductance state of neocortical neurons in vivo. *Nat. Rev. Neurosci.*, 4:739, 2003.

[19] T. W. Margrie, M. Brecht, and B. Sakmann. In vivo, low-resistance, whole-cell recordings from neurons in the anaesthetized and awake mammalian brain. *Pflug. Arch. Eur. J. Phy.*, 444:491, 2002.

[20] J. Waters and F. Helmchen. Background synaptic activity is sparse in neocortex. *J. Neurosci.*, 26:8267, 2006.

[21] M. D. McDonnell and L. M. Ward. The benefits of noise in neural systems: bridging theory and experiment. *Nat. Rev. Neurosci.*, 12:415, 2011.

[22] A. A. Faisal, L. P. J. Selen, and D. M. Wolpert. Noise in the nervous system. *Nat. Rev. Neurosci.*, 9:292, 2008.

[23] P. Dayan and L. F. Abbott. *Theoretical Neuroscience*. MIT Press, Cambridge MA, 2001.

[24] W. Gerstner, W. M. Kistler, R. Naud, and L. Paninski. *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press, 2014.

[25] U. Fano. Ionization Yield of Radiations. II. The Fluctuations of the Number of Ions. *Phys Rev*, 72(1):26–29, 1947.

[26] M. Churchland, M. Byron, J. Cunningham, L. P. Sugrue, G. S. Corrado M. R. Cohen, W. T. Newsome, A. M. Clark, P. Hosseini, B. B. Scott, D. C. Bradley, M. A. Smith, A. Kohn, J. A. Movshon, K. M. Armstrong, T. Moore, S. W. Chang, L. H. Snyder, S. G. Lisberger, N. J. Priebe, I. M. Finn, D. Ferster, S. I. Ryu, G. Santhanam, M. Sahani, and K. V. Shenoy. Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat. Neurosci.*, 13:369, 2010.

[27] A. Ponce-Alvarez, A. Thiele, T. D. Albright, G. R. Stoner, and G. Deco. Stimulus-dependent variability and noise correlations in cortical MT neurons. *PNAS*, 2013.

[28] C. W. Gardiner. *Handbook of Stochastic Methods*. Springer-Verlag, Berlin, 1985.

[29] E. D. Adrian. The impulses produced by sensory nerve endings. Part I. *J Physiol-London*, 61(1):49–72, 1926.

[30] H. K. Hartline and C. H. Graham. Nerve impulses from single receptors in the eye. *Journal of Cellular Physiology*, 1(2):277–295, 1932.

[31] C. Enroth-Cugell and J. G. Robson. The contrast sensitivity of retinal ganglion cells of the cat. *The Journal of Physiology*, 187(3):517–552, 1966.

[32] J. F. Brugge, D. J. Anderson, and J. E. Hind. Time structure of discharges in single auditory nerve fibers of the squirrel monkey in response to complex periodic sounds. *J. Neurophysiol.*, 32:386, 1969.

[33] D. R. Dvorak, L. G. Bishop, and H. E. Eckert. On the identification of movement detectors in the fly optic lobe. *J Comp Physiol*, 100(1):5–23, 1975.

[34] D. J. Simons. Response properties of vibrissa units in rat SI somatosensory neocortex. *Journal of Neurophysiology*, 41(3):798–820, 1978.

[35] R. Kubo. Statistical-mechanical theory of irreversible processes. i. general theory and simple applications to magnetic and conduction problems. *J. Phys. Soc. Jpn.*, 12:570, 1957.

[36] P. Hänggi and H. Thomas. Stochastic processes: Time evolution, symmetries and linear response. *Phys. Rep.*, 88:207, 1982.

[37] H. Risken. *The Fokker-Planck Equation*. Springer, Berlin, 1984.

[38] F. Rieke, D. Warland, R. de Ruyter van Steveninck, and W. Bialek. *Spikes: Exploring the neural code*. MIT Press, Cambridge, Massachusetts, 1996.

[39] H. Koendgen, C. Geisler, S. Fusi, X.-J. Wang, H.-R. Luescher, and M. Giugliano. The dynamical response properties of neocortical neurons to temporally modulated noisy inputs in vitro. *Cereb. Cortex*, 18:2086, 2008.

[40] B. W. Knight. Dynamics of encoding in a population of neurons. *J. Gen. Physiol.*, 59:734, 1972.

[41] B. Lindner and L. Schimansky-Geier. Transmission of noise coded versus additive signals through a neuronal ensemble. *Phys. Rev. Lett.*, 86:2934, 2001.

[42] N. Brunel, F. S. Chance, N. Fourcaud, and L. F. Abbott. Effects of synaptic noise and filtering on the frequency response of spiking neurons. *Phys. Rev. Lett.*, 86:2186, 2001.

[43] N. Fourcaud-Trocmé, D. Hansel, C. van Vreeswijk, and N. Brunel. How spike generation mechanisms determine the neuronal response to fluctuating inputs. *J. Neurosci.*, 23:11628, 2003.

[44] T. Tchumatchenko, A. Malyshev, F. Wolf, and M. Volgushev. Ultrafast population encoding by cortical neurons. *J. Neurosci.*, 31:12171, 2011.

[45] N. Brunel. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J. Comput. Neurosci.*, 8:183, 2000.

[46] P. A. Franken and J. F. Ward. Optical Harmonics and Nonlinear Phenomena. *Rev. Mod. Phys.*, 35(1):23–39, 1963.

[47] E. Hendry, P. J. Hale, J. Moger, A. K. Savchenko, and S. A. Mikhailov. Coherent Nonlinear Optical Response of Graphene. *Phys. Rev. Lett.*, 105(9):097401–4, 2010.

[48] W. R. Zipfel, R. M. Williams, R. Christie, A. Y. Nikitin, B. T. Hyman, and W. W. Webb. Live tissue intrinsic emission microscopy using multiphoton-excited native fluorescence and second harmonic generation. *Proc. Natl. Acad. Sci.*, 100(12):7075–80, 2003.

[49] B. Gleich and J. Weizenecker. Tomographic imaging using the nonlinear response of magnetic particles. *Nature*, 435(7046):1214–7, 2005.

[50] P. Z. Marmarelis and K. I. Naka. White-Noise Analysis of a Neuron Chain: An Application of the Wiener Theory. *Science*, 175(4027):1276–8, 1972.

[51] Y. Yu, R. Romero, and T. S. Lee. Preference of Sensory Neural Coding for 1/f Signals. *Physical review letters*, 94(10):108103–4, 2005.

[52] A. N. Temchin, A. Recio-Spinoso, P. van Dijk, and M. A. Ruggero. Wiener kernels of chinchilla auditory-nerve fibers: Verification using responses to tones, clicks, and noise and comparison with basilar-membrane vibrations. *J. Neurophysiol.*, 93(6):3635–48, 2005.

[53] A. Recio-Spinoso, A. N. Temchin, P. van Dijk, Y. H. Fan, and M. A. Ruggero. Wiener-kernel analysis of responses to noise of chinchilla auditory-nerve fibers. *J. Neurophysiol.*, 93(6):3615–34, 2005.

[54] N. Brunel and V. Hakim. Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Comput.*, 11:1621, 1999.

[55] M. Deger, M. Helias, S. Cardanobile, F. M. Atay, and S. Rotter. Nonequilibrium dynamics of stochastic point processes with refractoriness. *Phys. Rev. E.*, 82:021129, 2010.

[56] W. Bialek, M. Deweese, F. Rieke, and D. Warland. Bits and brains - information-flow in the nervous-system. *Physica A.*, 200:581, 1993.

[57] R. Shannon. The mathematical theory of communication. *Bell. Syst. Tech. J.*, 27:379, 1948.

[58] R. B. Ash. *Information Theory.* Dover Publications Inc., New York, 1990.

[59] R.R. De Ruyter van Steveninck, G. D. Lewen, S. P. Strong, R. Koberle, and W. Bialek. Reproducibility and variability in neural spike trains. *Science*, 275:1805, 1997.

[60] G. T. Buracas, A. M. Zador, M. R. Deweese, and T. D. Albright. Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. *Neuron*, 20(5):959–969, 1998.

[61] S. P. Strong, R. Koberle, R. R. D. van Steveninck, and W. Bialek. Entropy and information in neural spike trains. *Phys. Rev. Lett.*, 80:197, 1998.

[62] A. Borst and F. Theunissen. Information theory and neural coding. *Nat. Neurosci.*, 2:947, 1999.

[63] F. Gabbiani. Coding of time-varying signals in spike trains of linear and half-wave rectifying neurons. *Network: Computation in Neural Systems*, 7:61, 1996.

[64] R. Krahe and F. Gabbiani. Burst firing in sensory systems. *Nat. Rev. Neurosci.*, 5:13, 2004.

[65] Z. N. Aldworth, A. G. Dimitrov, G. I. Cummins, T. Gedeon, and J. P. Miller. Temporal encoding in a nervous system. *PLoS Computational Biology*, 2011.

[66] D. Bernardi and B. Lindner. A frequency-resolved mutual information rate and its application to neural systems. *J Neurophysiol*, 113(5):1342–1357, 2015.

[67] J. Laudanski, S. Coombes, A. R. Palmer, and C. J. Sumner. Mode-Locked Spike Trains in Responses of Ventral Cochlear Nucleus Chopper and Onset Neurons to Periodic Stimuli. *J. Neurophysiol.*, 103(3):1226–37, 2010.

[68] L. Lapicque. Recherches quantitatives sur l'excitation électrique des nerfs traitéecomme une polarization. *J. Physiol. Pathol. Gen.*, 9:620, 1907.

[69] N. Brunel and M. C. W. van Rossum. Lapicque's 1907 paper: from frogs to integrate-and-fire. *Biol Cybern*, 2007.

[70] A. N. Burkitt. A review of the integrate-and-fire neuron model: I. homogeneous synaptic input. *Biol. Cyber.*, 95:1, 2006.

[71] A. N. Burkitt. A review of the integrate-and-fire neuron model: II. inhomogeneous synaptic input and network properties. *Biol. Cyber.*, 95:97, 2006.

[72] N. Fourcaud and N. Brunel. Dynamics of the firing probability of noisy integrate-and-fire neurons. *Neural Comput.*, 14:2057, 2002.

[73] A. J. F. Siegert. On the first passage time problem. *Phys. Rev.*, 81:617, 1951.

[74] L. M. Ricciardi. *Diffusion Processes and Related Topics on Biology.* Springer-Verlag, Berlin, 1977.

[75] B. Lindner, L. Schimansky-Geier, and A. Longtin. Maximizing spike train coherence or incoherence in the leaky integrate-and-fire model. *Phys. Rev. E.*, 66:031916, 2002.

[76] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions.* Dover, New York, 1970.

[77] C. Morris and H. Lecar. Voltage oscillations in the barnacle giant muscle fiber. *Biophys. J.*, 35:193, 1981.

[78] P. J. Thomas and B. Lindner. Asymptotic Phase for Stochastic Oscillators. *Physical Review letters*, 113(25):254101–5, 2014.

[79] P. M. Morse and H. Feshbach. *Methods of Theoretical Physics.* McGraw-Hill, New York, Toronto, London, 1953.

[80] H. Schulz. *Physik mit Bleistift.* Verlag Harri Deutsch, Frankfurt, 2009.

[81] W. E. Boyce and R. C. DiPrima. *Elementary differential equations and boundary value problems.* Wiley New York, 5 edition, 1992.

[82] B. Lindner. *Coherence and Stochastic Resonance in Nonlinear Dynamical Systems.* Logos-Verlag, Berlin, 2002.

[83] J. Schuecker, M. Diesmann, and M. Helias. Modulated escape from a metastable state driven by colored noise. *Phys Rev E*, 92(5):052119–11, 2015.

[84] R. D. Vilela and B. Lindner. Are the input parameters of white-noise-driven integrate & fire neurons uniquely determined by rate and CV? *J. Theor. Biol.*, 257:90, 2009.

[85] S. Ostojic and N. Brunel. From spiking neuron models to linear-nonlinear models. *PLoS Comput. Biol.*, 7:e1001056, 2011.

[86] J. P. Keener, F. C. Hoppensteadt, and J. Rinzel. Integrate-and-Fire Models of Nerve Membrane Response to Oscillatory Input. *SIAM J. Appl. Math.*, 41(3):503–17, 1981.

[87] J. A. Freund, A. Nikitin, and N. G. Stocks. Phase locking below rate threshold in noisy model neurons. *Neural Comput.*, 22:599, 2010.

[88] T Tateno and Y Jimbo. Stochastic mode-locking for a noisy integrate-and-fire oscillator. *Phys Lett A*, 271(4):227–36, 2000.

[89] D. R. Chialvo, A. Longtin, and J. Müller-Gerking. Stochastic resonance in models of neuronal ensembles. *Phys. Rev. E.*, 55:1798, 1997.

[90] T. Cover and J. Thomas. *Elements of Information Theory.* Wiley, New-York, 1991.

[91] G. Silberberg, M. Bethge, H. Markram, and K. Pawelzik und M. Tsodyks. Dynamics of population rate codes in ensembles of neocortical neurons. *J. Neurophysiol.*, 91:704, 2004.

[92] I. M. Ryzhik and I. S. Gradshtein. *Tables of series, products, and integrals.* VEB Deutscher Verlag der Wissenschaften, 1963.

[93] J. Lagarias. Euler's constant: Euler's work and modern developments. *Bulletin of the American Mathematical Society*, 50(4):527–628, 2013.

[94] D. R. Cox. *Renewal Theory.* Methuen, London, 1962.

[95] D. R. Cox and P. A. W. Lewis. *The Statistical Analysis of Series of Events.* Chapman and Hall, London, 1966.

[96] P. Reimann, C. Van den Broeck, H. Linke, P. Hänggi, M. Rubi, and A. Perez-Madrid. Giant acceleration of free diffusion by use of tilted periodic potentials. *Phys. Rev. Lett.*, 87:010602, 2001.

[97] B. Lindner and I. M. Sokolov. Giant diffusion of underdamped particles in a biased periodic potential. *Physical Review E*, 93(4):042106, 2016.

[98] N. Brunel and J. P. Nadal. Mutual information, fisher information, and population coding. *Neural Comput.*, 10:1731, 1998.

[99] B. Lindner, M. J. Chacron, and A. Longtin. Integrate-and-fire neurons with threshold noise - a tractable model of how interspike interval correlations affect neuronal signal transmission. *Phys. Rev. E.*, 72:021911, 2005.

[100] S O Voronenko, W Stannat, and Benjamin Lindner. Shifting Spike Times or Adding and Deleting Spikes—How Different Types of Noise Shape Signal Transmission in Neural Populations. *JMN*, 5(1), 2015.

[101] D. K. Warland, P. Reinagel, and M. Meister. Decoding visual information from a population of retinal ganglion cells. *J. Neurophysiol.*, 78:2336, 1997.

[102] O. Marre, V. Botella-Soler, K. D. Simmons, T. Mora, G. Tkačik, and M. J. Berry II. High accuracy decoding of dynamical motion from a large retinal population. *PLoS computational biology*, 11(7):e1004304, 2015.

[103] R. Q. Quiroga and S. Panzeri. Extracting information from neuronal populations: information theory and decoding approaches. *Nature Review Neuroscience*, 10:173, 2009.

[104] A. B. Neiman and D. F. Russell. Sensory coding in oscillatory electroreceptors of paddlefish. *Chaos*, 21:047505, 2011.

[105] L. Glass. Synchronization and rhythmic processes in physiology. *Nature*, 410(6825):277–284, 2001.

[106] A. Neiman, L. Schimansky-Geier, F. Moss, B. Shulgin, , and J. J. Collins. Synchronization of noisy systems by stochastic signals. *Phys. Rev. E.*, 60:284, 1999.

[107] J. A. Freund, L. Schimansky-Geier, and P. Hänggi. Frequency and phase synchronization in stochastic systems. *Chaos*, 13:225, 2003.

[108] H. Spekreijse and H. Oosting. Linearizing: a method for analysing and synthesizing nonlinear systems. *Biological Cybernetics*, 7(1):22–31, 1970.

[109] M. Carandini, F. Mechler, C. S. Leonard, and J. A. Movshon. Spike train encoding by regular-spiking cells of the visual cortex. *Journal of Neurophysiology*, 76(5):3425–3441, 1996.

[110] A. K. Lee, I. D. Manns, B. Sakmann, and M. Brecht. Whole-cell recordings in freely moving rats. *Neuron*, 51:399, 2006.

[111] György Buzsáki and Xiao-Jing Wang. Mechanisms of Gamma Oscillations. *Annu Rev Neurosci*, 35(1):203–225, July 2012.

[112] M. Stemmler. A single spike suffices: the simplest form of stochastic resonance in neuron models. *Network*, 7:687, 1996.

[113] C. L. Nikias and A. P. Petropulu. *Higher-order spectral analysis*. PTR Prentice Hall, 1993.

[114] Y. Dan, J. M. Alonso, W. M. Usrey, and R. C. Reid. Coding of visual information by precisely correlated spikes in the lateral geniculate nucleus. *Nature neuroscience*, 1(6), 1998.

[115] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440:1007, 2006.

[116] I. E. Ohiorhenuan, F. Mechler, K. P. Purpura, A. M. Schmid, Q. Hu, and J. D. Victor. Sparse coding and high-order correlations in fine-scale cortical networks. *Nature*, 466:617, 2010.

[117] M. J. E. Richardson. Firing-rate response of linear and nonlinear integrate-and-fire neurons to modulated current-based and conductance-based synaptic drive. *Phys. Rev. E.*, 76:021919, 2007.

[118] M. Helias, M. Deger, S. Rotter, and M. Diesmann. Instantaneous Non-Linear Processing by Pulse-Coupled Threshold Units. *PLoS Comput. Biol.*, 6(9), 2010.

[119] M. Helias, M. Deger, S. Rotter, and M. Diesmann. Finite post synaptic potentials cause a fast neuronal response. *Frontiers in neuroscience*, 5, 2011.

Publications concerning this thesis

Articles in scientific journals

- S.O. Voronenko and B. Lindner: **Weakly nonlinear response of noisy neurons**, *New Journal of Physics* **19** (2017) 033038
- S.O. Voronenko and B. Lindner: **Nonlinear information transmission by the neural count**, *in preparation*

Selbständigkeitserklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den 6. Februar 2018 Sergej Voronenko

Danksagung

Diese Arbeit wäre ohne meinen Betreuer, Prof. Benjamin Lindner, und dessen herausragende Betreuung niemals möglich gewesen. Über die letzten Jahre war er mir nicht nur mein wichtigster wissenschaftlicher Lehrer, der in mir die Begeisterung für analytisches Rechnen weckte, sondern auch ein wertvoller Mentor, der mich lehrte nicht vor schwierigen Aufgaben zurückzuschrecken. Dafür bleibe ich für immer dankbar.

Großer Dank gebührt auch allen Mitglieder der AG "Theorie komplexer Systeme und Neurophysik", die meine Zeit am Institut stets zu einem angenehmen Erlebnis gemacht haben. Dafür danke ich Euch!

Besonderer Dank gilt all denen, die beim Korrekturlesen dieser Arbeit mitgewirkt haben. Insbesondere möchte ich Davide Bernardi, Alexander van Meegen, Philip Lauinger und Thomas Schlieter danken. Vielen Dank für Eure Mühe!

Zuletzt möchte ich mich bei meiner Ehefrau, Beatrice Voronenko, bedanken. Den Wert deiner Unterstützung und deines Zuspruches in den letzten Zügen der Fertigstellung dieser Arbeit kann man nicht in Worte fassen. Danke dafür, dass es Dich gibt!